Lectures for the 27th IAU ISYA
Ifrane, 2nd - 23rd July 2004

Statistical Astronomy

Martin Hendry,
Dept of Physics and Astronomy
University of Glasgow, UK

UNIVERSITY of GLASGOW

IAU

http://www.astro.gla.ac.uk/users/martin/isya/

Parameter estimation: the Gaussian approximation

Posterior Likelihood Prior

$$\boxed{p(\text{model} \mid \text{data}, I) \propto p(\text{data} \mid \text{model}, I) \times p(\text{model} \mid I)}$$

UNIVERSITY of GLASGOW

Parameter estimation: the Gaussian approximation

$$\boxed{p(\theta \mid \text{data}, I) \propto p(\text{data} \mid \theta, I) \times p(\theta \mid I)}$$

'Best' estimator: $\left.\frac{\partial p(\theta \mid \text{data}, I)}{\partial \theta}\right|_{\theta=\theta_0} = 0$ Maximise posterior likelihood

Equivalently, we can define $\ell = \log p(\theta \mid \text{data}, I)$ and compute $\left.\frac{\partial \ell}{\partial \theta}\right|_{\theta=\theta_0} = 0$

Taylor expand $\ell(\theta)$ around $\theta=\theta_0$:

$$\boxed{\ell(\theta) = \ell(\theta_0) + \cancel{\left.\frac{\partial \ell}{\partial \theta}\right|_{\theta=\theta_0}(\theta-\theta_0)} + \frac{1}{2}\left.\frac{\partial^2 \ell}{\partial \theta^2}\right|_{\theta=\theta_0}(\theta-\theta_0)^2 + \ldots}$$

Parameter estimation: the Gaussian approximation

$$p(\theta \mid \text{data}, \ I) = \exp\left[\ell(\theta)\right]$$

Neglecting higher order terms in $\ell(\theta)$ ← Gaussian approximation

$$p(\theta \mid \text{data}, \ I) \propto \exp\left(-\frac{A}{2}(\theta - \theta_0)^2\right)$$

where $A = -\left.\dfrac{\partial^2 \ell}{\partial \theta^2}\right|_{\theta=\theta_0}$

This is equivalent to a normal distribution, with $\sigma^{-2} = A = -\left.\dfrac{\partial^2 \ell}{\partial \theta^2}\right|_{\theta=\theta_0}$

Can summarise inference from posterior by

$$\boxed{\theta = \theta_0 \pm \sigma}$$

UNIVERSITY *of* GLASGOW

Parameter estimation: 2-D case

Recall our definition of *variance*

$$\mathrm{var}[x] \;=\; \int_{-\infty}^{\infty} \left(x - \langle x \rangle\right)^2 p(x \mid I)\,dx$$

Extend to 2 variables - *covariance*

$$\mathrm{cov}[x, y] \;=\; \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left(x - \langle x \rangle\right)\left(y - \langle y \rangle\right) p(x, y \mid I)\,dx\,dy$$

If x and y are independent, $\mathrm{cov}[x, y] \;=\; 0$

This is because $p(x, y \mid I) = p(x \mid I)\, p(y \mid I)$

UNIVERSITY of GLASGOW

Parameter estimation: 2-D case

$$\boxed{p(\theta_1, \theta_2 \mid \text{data}, \; I) \quad \propto \quad p(\text{data} \;\mid \theta_1, \theta_2, I) \times p(\theta_1, \theta_2 \mid I)}$$

'Best' estimator: $$\left.\frac{\partial p(\theta_1, \theta_2 \mid \text{data}, I)}{\partial \theta_j}\right|_{\theta_j = \theta_{0\,j}} = 0$$

Compute $$\left.\frac{\partial \ell}{\partial \theta_j}\right|_{\theta_j = \theta_{0\,j}} = 0$$ where $\ell = \log p(\theta_1, \theta_2 \mid \text{data}, I)$

UNIVERSITY of GLASGOW

Parameter estimation: 2-D case

Taylor expand $\ell(\theta_1, \theta_2)$ around θ_{0j} :

$$\boxed{\begin{aligned}\ell(\theta_1,\theta_2) \;=\; & \ell(\theta_{01},\theta_{02}) + \cancel{\left.\frac{\partial \ell}{\partial \theta_1}\right|_{\theta_j=\theta_{0j}} (\theta_1 - \theta_{01})} + \cancel{\left.\frac{\partial \ell}{\partial \theta_2}\right|_{\theta_j=\theta_{0j}} (\theta_2 - \theta_{02})} + \\ & \frac{1}{2}\left[\left.\frac{\partial^2 \ell}{\partial \theta_1^{\,2}}\right|_{\theta_j=\theta_{0j}} (\theta_1-\theta_{01})^2 + \left.\frac{\partial^2 \ell}{\partial \theta_2^{\,2}}\right|_{\theta_j=\theta_{0j}} (\theta_2-\theta_{02})^2 + 2\left.\frac{\partial^2 \ell}{\partial \theta_1 \partial \theta_2}\right|_{\theta_j=\theta_{0j}} (\theta_1-\theta_{01})(\theta_2-\theta_{02}) \right] + \ldots\end{aligned}}$$

$$p(\theta_1, \theta_2 \mid \text{data}, \; I) \;\propto\; \exp\left[\ell(\theta_1,\theta_2)\right]$$

$$\propto\; \exp\left[-\tfrac{1}{2} Q\right]$$

← Gaussian approximation

χ^2

Maximising likelihood
$\equiv$ Minimising χ^2

UNIVERSITY *of* GLASGOW

Parameter estimation: 2-D case

Taylor expand $\ell(\theta_1, \theta_2)$ around θ_{0j} :

$$\boxed{Q = \begin{pmatrix} \theta_1 - \theta_{10} & \theta_2 - \theta_{20} \end{pmatrix} \begin{bmatrix} A & C \\ C & B \end{bmatrix} \begin{pmatrix} \theta_1 - \theta_{10} \\ \theta_2 - \theta_{20} \end{pmatrix}}$$

Quadratic form

where $A = \left.\dfrac{\partial^2 \ell}{\partial \theta_1{}^2}\right|_{\theta_j = \theta_{0j}}$ $\quad B = \left.\dfrac{\partial^2 \ell}{\partial \theta_2{}^2}\right|_{\theta_j = \theta_{0j}}$ $\quad C = \left.\dfrac{\partial^2 \ell}{\partial \theta_1 \partial \theta_2}\right|_{\theta_j = \theta_{0j}}$

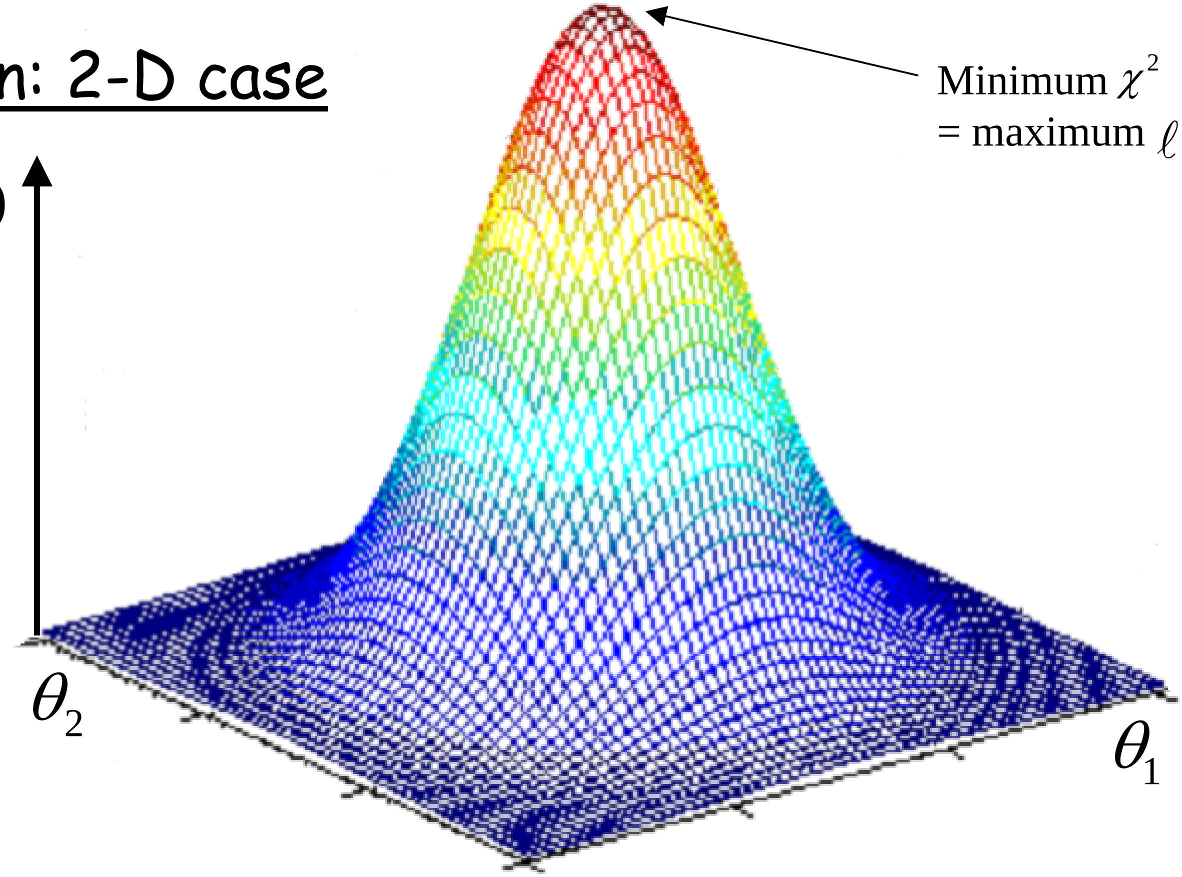
This is a bivariate normal distribution with covariance matrix

$$\sigma_{ij}^2 = \mathrm{cov}_{ij} = \left\langle (\theta_i - \theta_{i0})(\theta_j - \theta_{j0}) \right\rangle = \left[-\frac{\partial^2 \ell}{\partial \theta_i \partial \theta_j} \right]^{-1}$$

Fisher information matrix

UNIVERSITY of GLASGOW

Parameter estimation: 2-D case



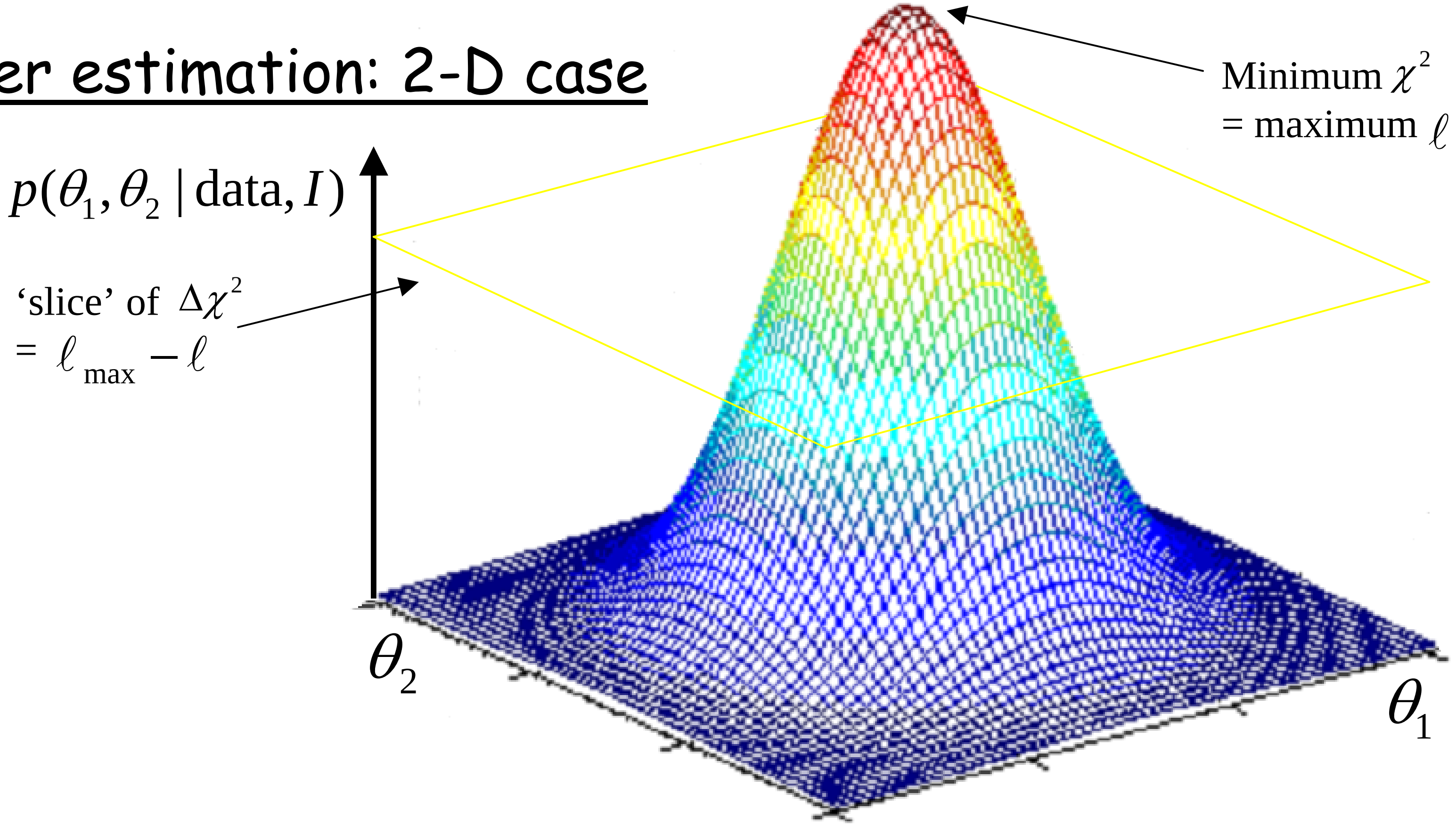
This is a bivariate normal distribution with covariance matrix

$$\sigma_{ij}^2 \quad = \quad \mathrm{cov}_{ij} \quad = \quad \left\langle \left(\theta_i - \theta_{i0}\right)\left(\theta_j - \theta_{j0}\right) \right\rangle \quad = \quad \left[-\frac{\partial^2 \ell}{\partial \theta_i \partial \theta_j} \right]^{-1}$$

Fisher information matrix

UNIVERSITY of GLASGOW

Parameter estimation: 2-D case



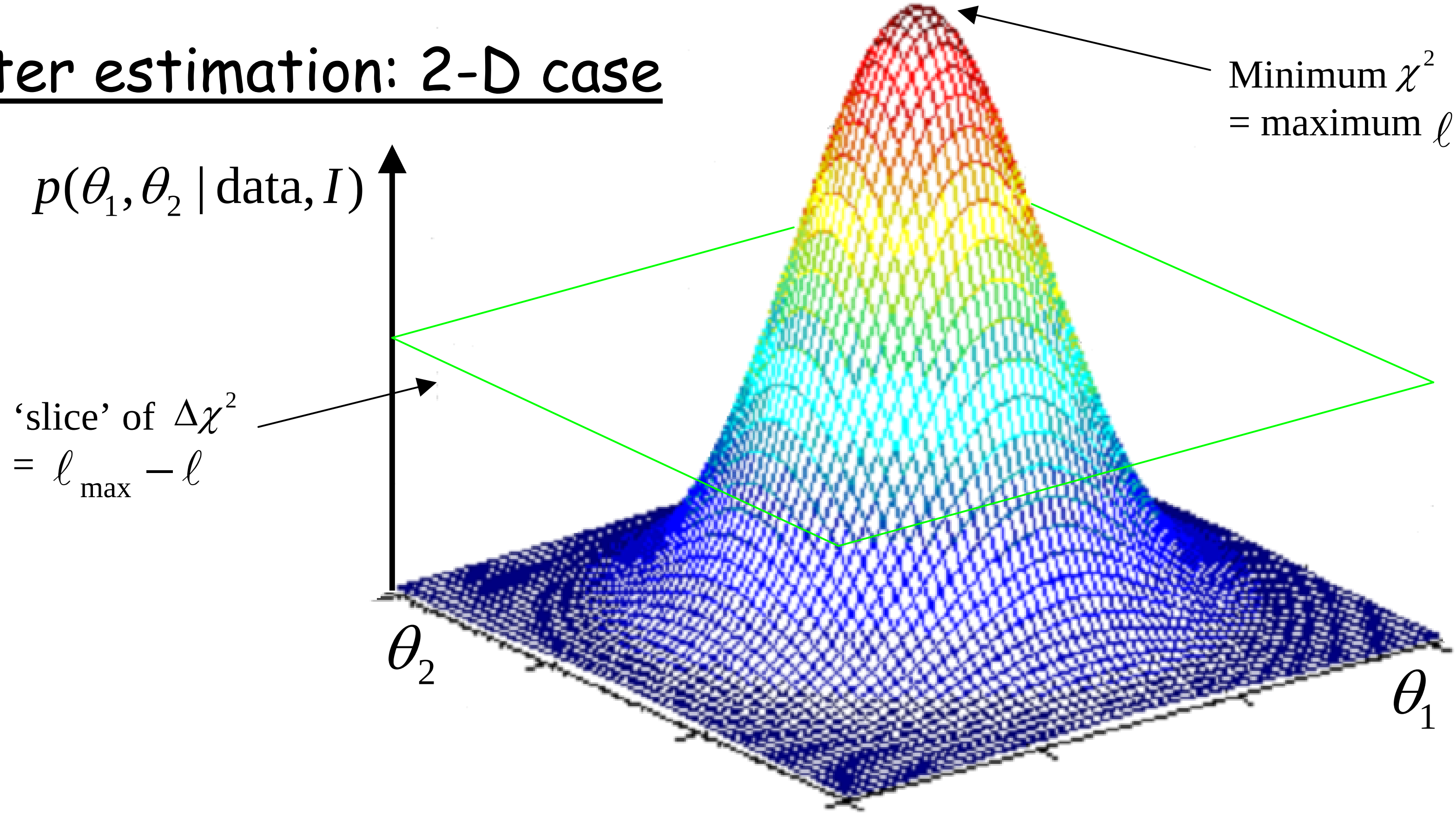
This is a bivariate normal distribution with covariance matrix

$$\sigma_{ij}^2 = \mathrm{cov}_{ij} = \left\langle (\theta_i - \theta_{i0})(\theta_j - \theta_{j0}) \right\rangle = \left[-\frac{\partial^2 \ell}{\partial \theta_i \partial \theta_j} \right]^{-1}$$

Fisher information matrix

UNIVERSITY of GLASGOW

Parameter estimation: 2-D case



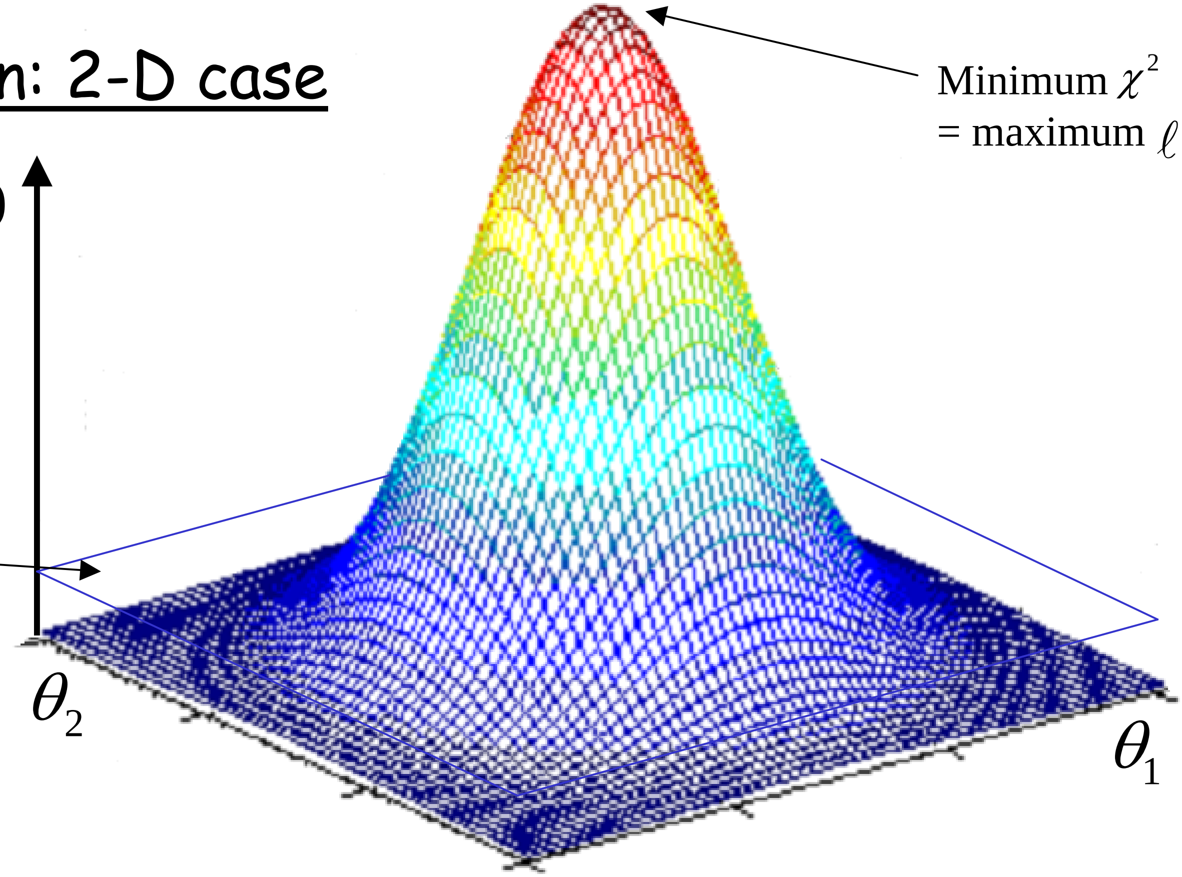
This is a bivariate normal distribution with covariance matrix

$$\sigma_{ij}^2 = \text{cov}_{ij} = \left\langle (\theta_i - \theta_{i0})(\theta_j - \theta_{j0}) \right\rangle = \left[-\frac{\partial^2 \ell}{\partial\theta_i \partial\theta_j} \right]^{-1}$$

Fisher information matrix

UNIVERSITY of GLASGOW

Parameter estimation: 2-D case



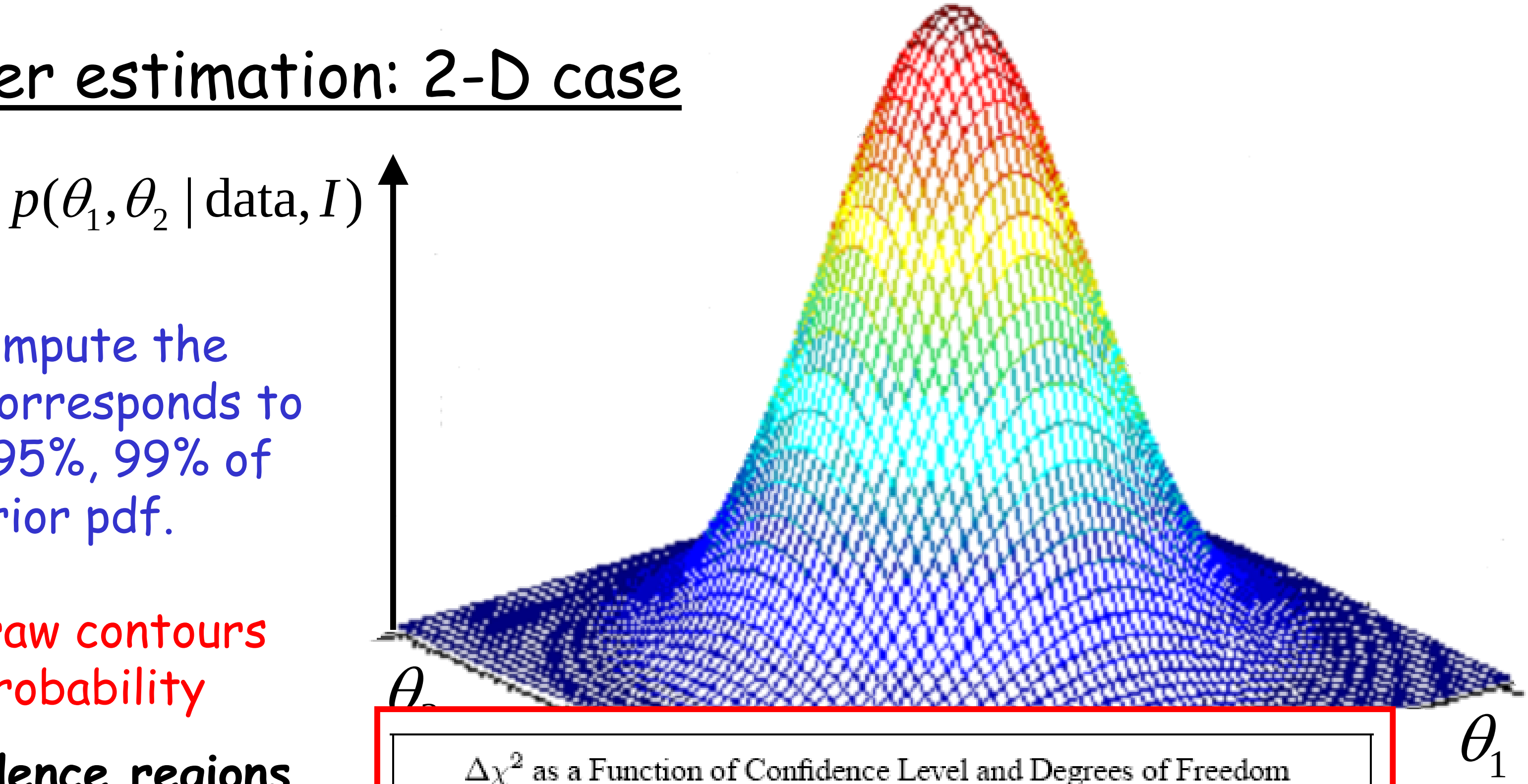
This is a bivariate normal distribution with covariance matrix

$$\sigma_{ij}^2 \quad = \quad \mathrm{cov}_{ij} \quad = \quad \left\langle (\theta_i - \theta_{i0})(\theta_j - \theta_{j0}) \right\rangle \quad = \quad \left[-\frac{\partial^2 \ell}{\partial \theta_i \partial \theta_j} \right]^{-1}$$

Fisher information matrix

UNIVERSITY *of* GLASGOW

Parameter estimation: 2-D case



We can compute the $\Delta\chi^2$ that corresponds to e.g. 68%, 95%, 99% of the posterior pdf.

We can draw contours of equal probability

$\Rightarrow$ **Confidence regions for the parameters**

Extends easily to N parameters – or *degrees of freedom*

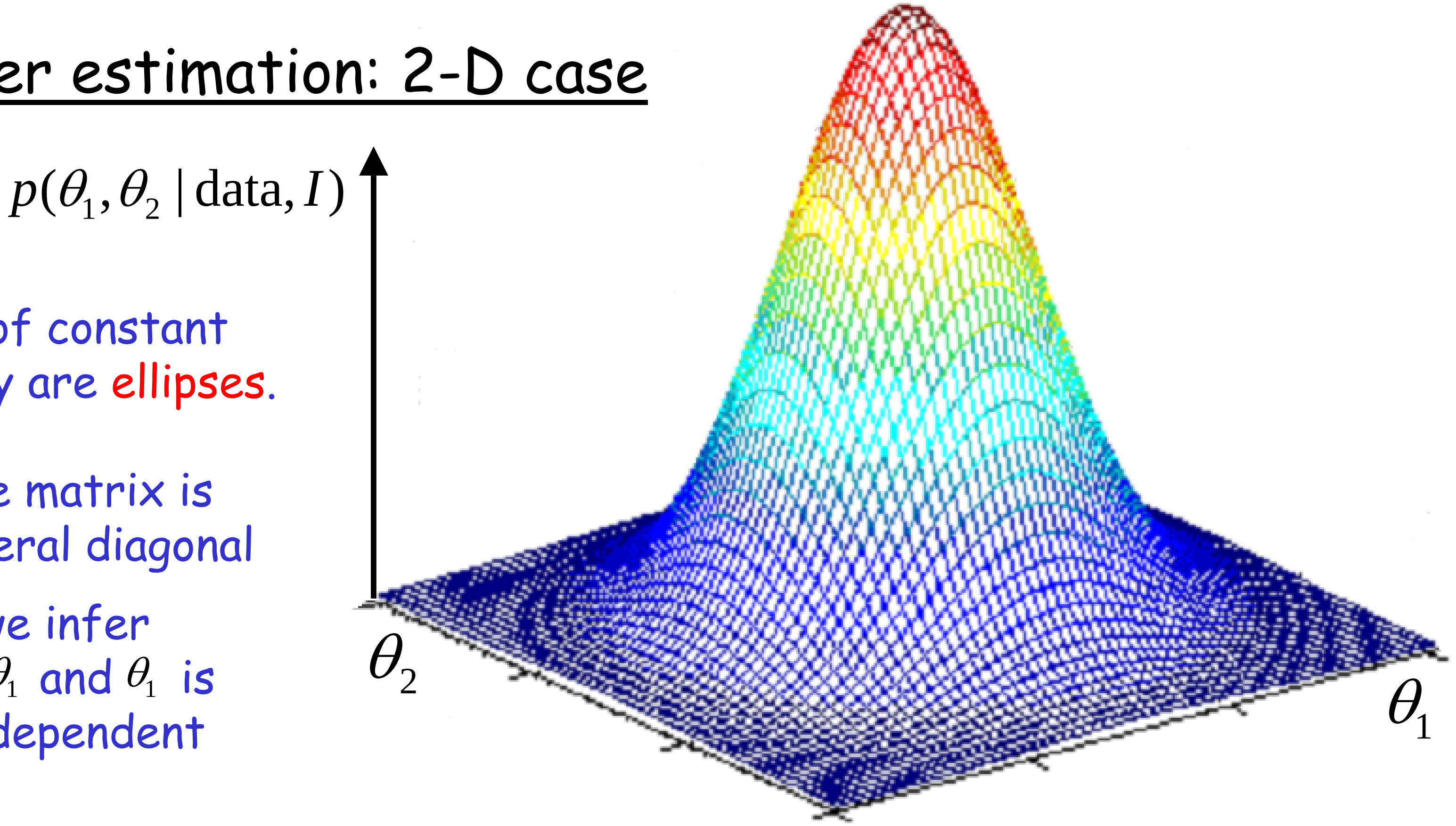
$\Delta\chi^2$ as a Function of Confidence Level and Degrees of Freedom

	ν					
p	1	2	3	4	5	6
68.3%	1.00	2.30	3.53	4.72	5.89	7.04
90%	2.71	4.61	6.25	7.78	9.24	10.6
95.4%	4.00	6.17	8.02	9.70	11.3	12.8
99%	6.63	9.21	11.3	13.3	15.1	16.8
99.73%	9.00	11.8	14.2	16.3	18.2	20.1
99.99%	15.1	18.4	21.1	23.5	25.7	27.8

From Numerical Recipes

UNIVERSITY of GLASGOW

Parameter estimation: 2-D case



Contours of constant probability are ellipses.

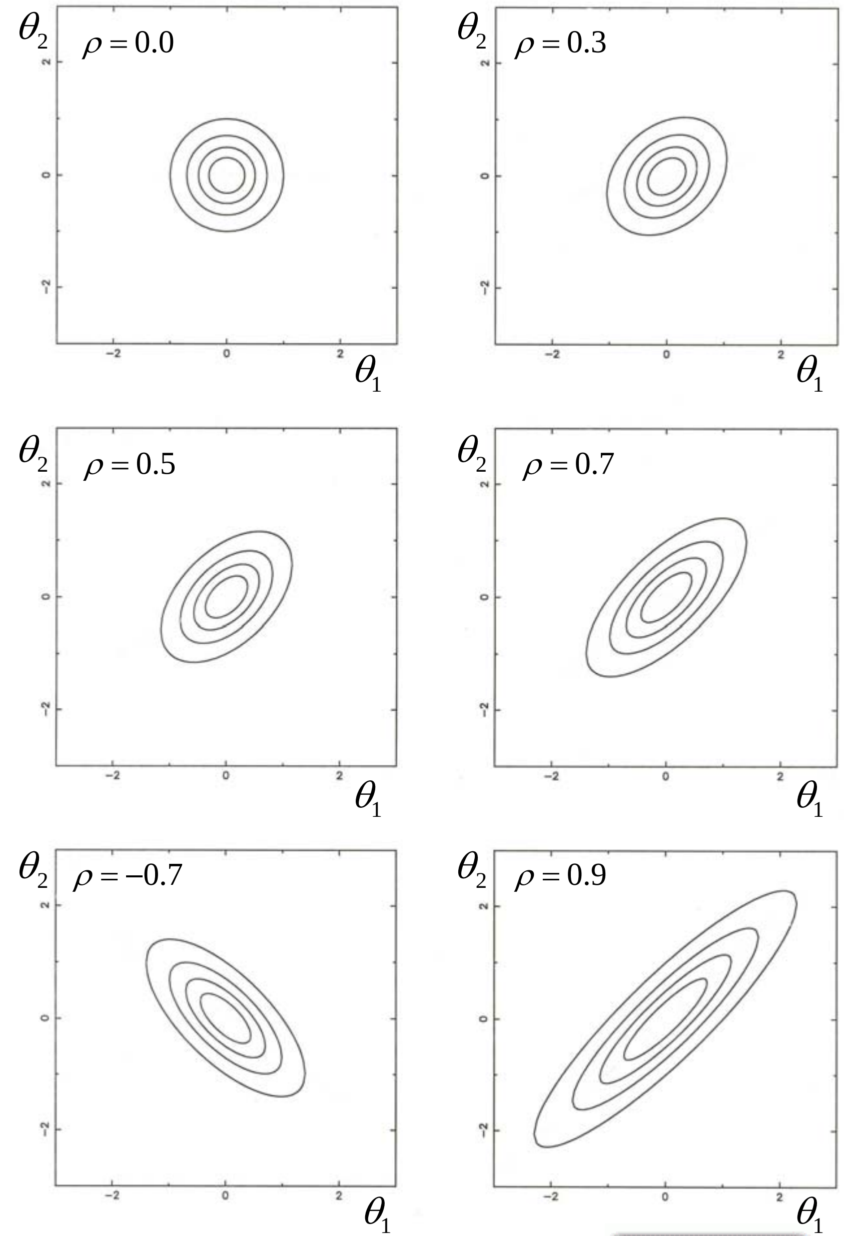
Covariance matrix is *not* in general diagonal

$\Rightarrow$ What we infer about θ_1 and θ_1 is *not* independent

Parameter estimation: 2-D case

Can define *correlation coefficient*

$$\rho = \frac{\mathrm{cov}[\theta_1, \theta_2]}{\sqrt{\mathrm{var}[\theta_1]}\,\sqrt{\mathrm{var}[\theta_2]}} \qquad -1 \le \rho \le 1$$



Covariance matrix becomes less diagonal

$\Rightarrow$ $|\rho|$ increases

$\Rightarrow$ isoprobability contours elongate

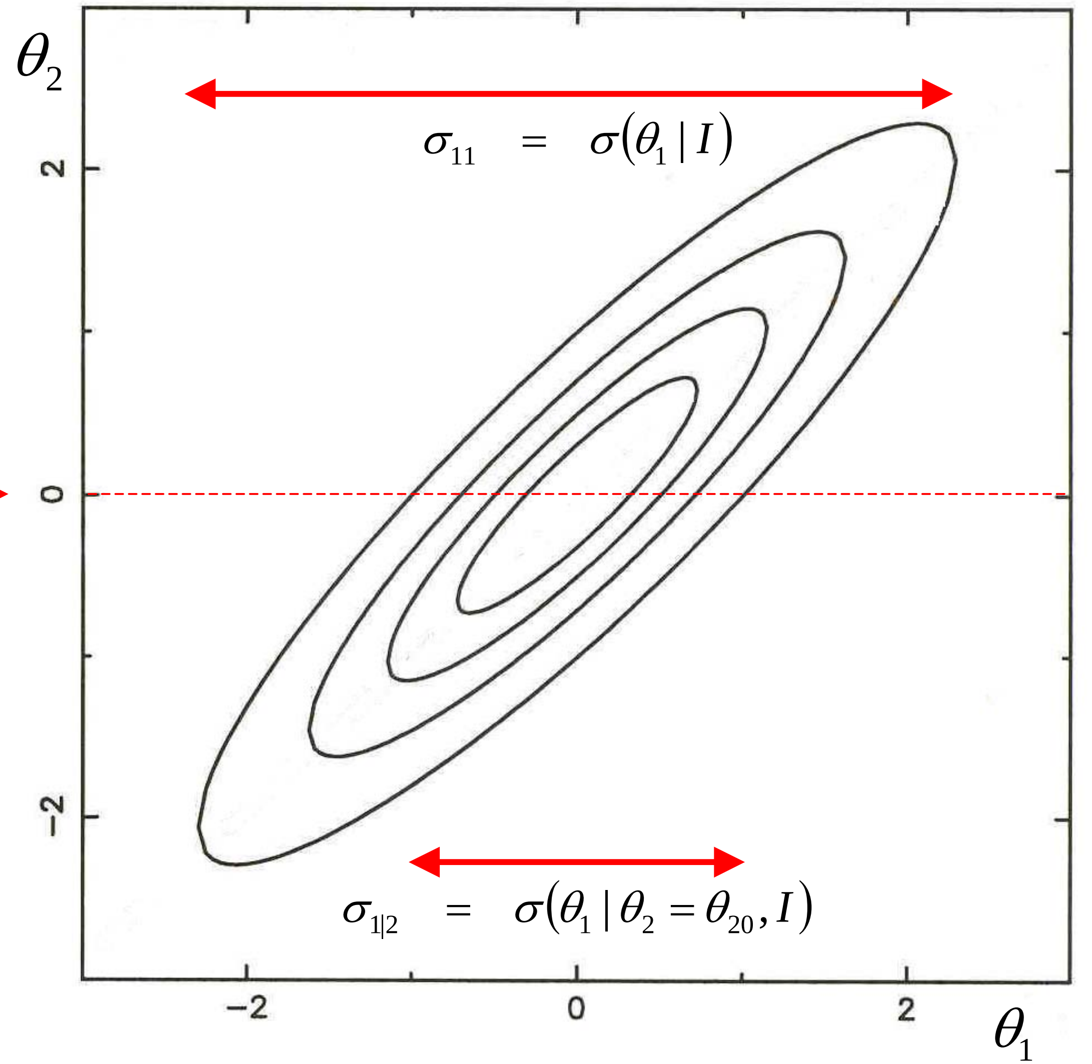
Very important if we are interested only in *one* parameter

UNIVERSITY of GLASGOW

Parameter estimation: 2-D case

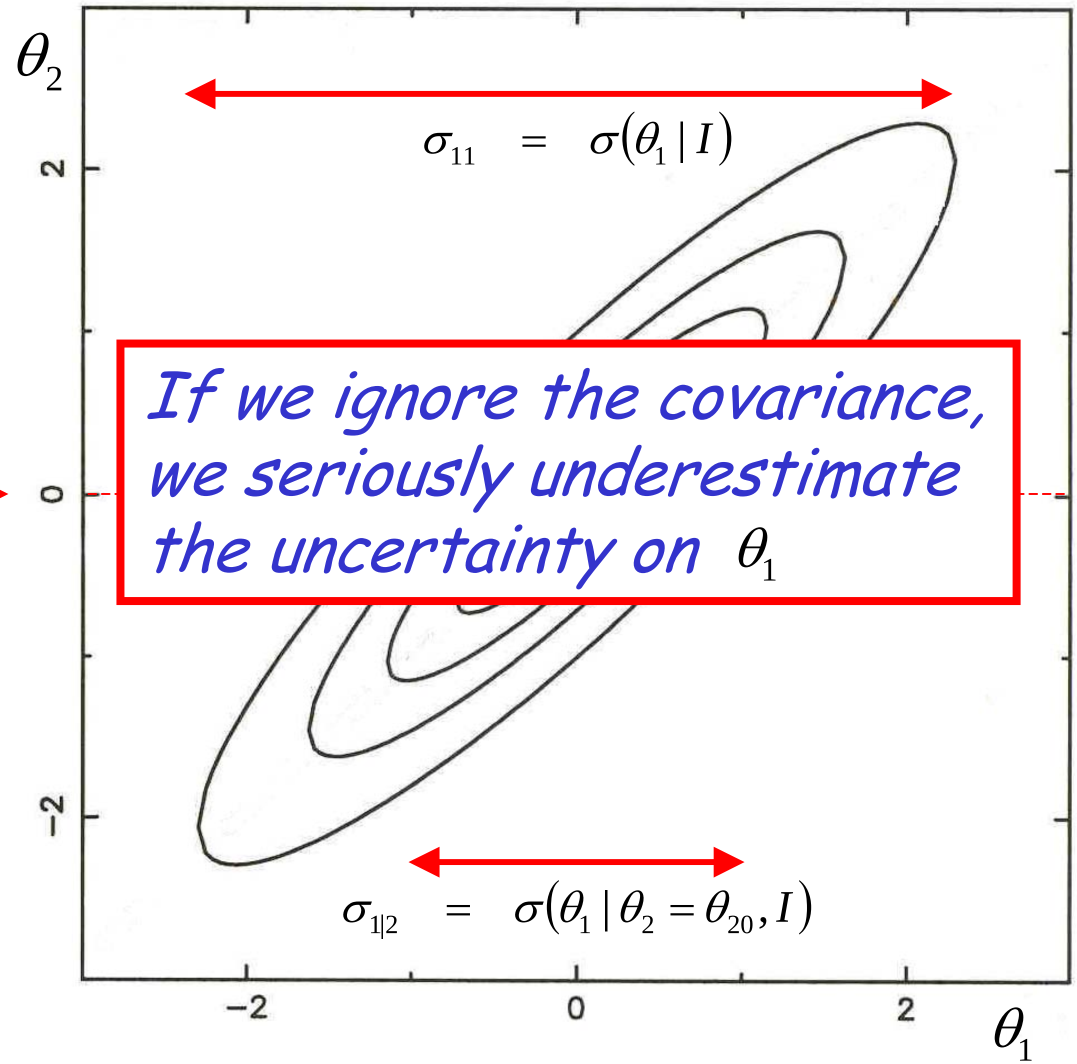
'Best-fit' value of θ_2 , found from

$$\left.\frac{\partial \ell}{\partial \theta_j}\right|_{\theta_j=\theta_{0j}} = 0$$

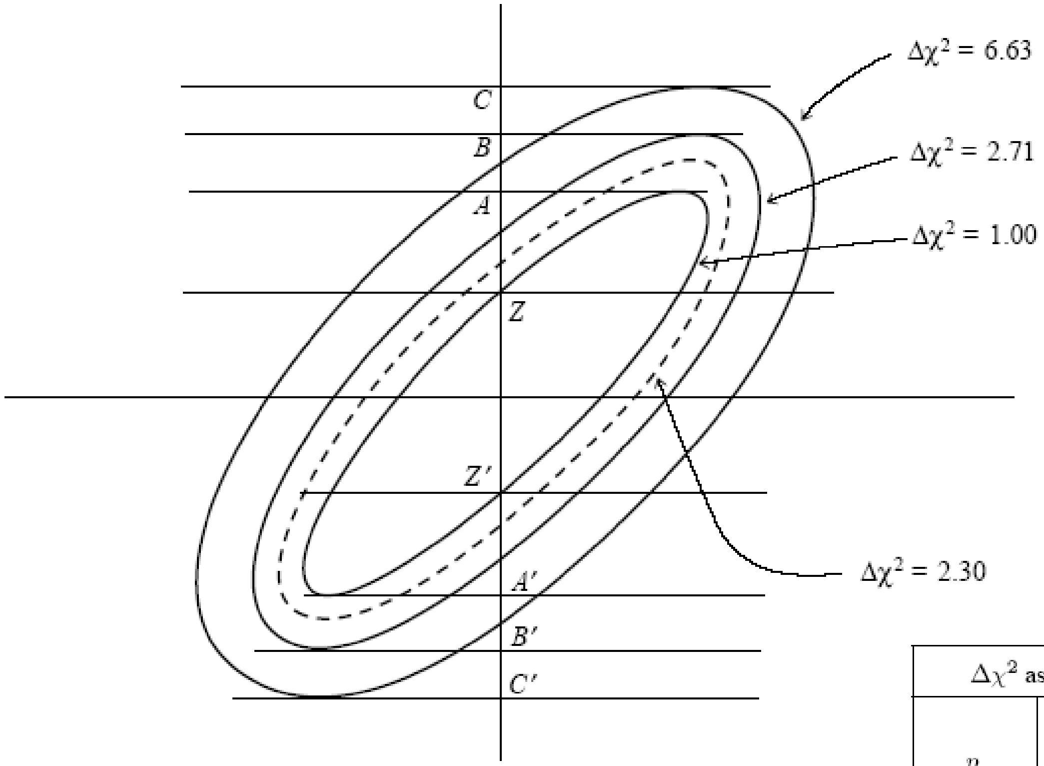


Parameter estimation: 2-D case

'Best-fit' value of θ_2 , found from $\left.\frac{\partial \ell}{\partial \theta_j}\right|_{\theta_j=\theta_{0j}} = 0$



Parameter estimation: 2-D case



From Numerical Recipes

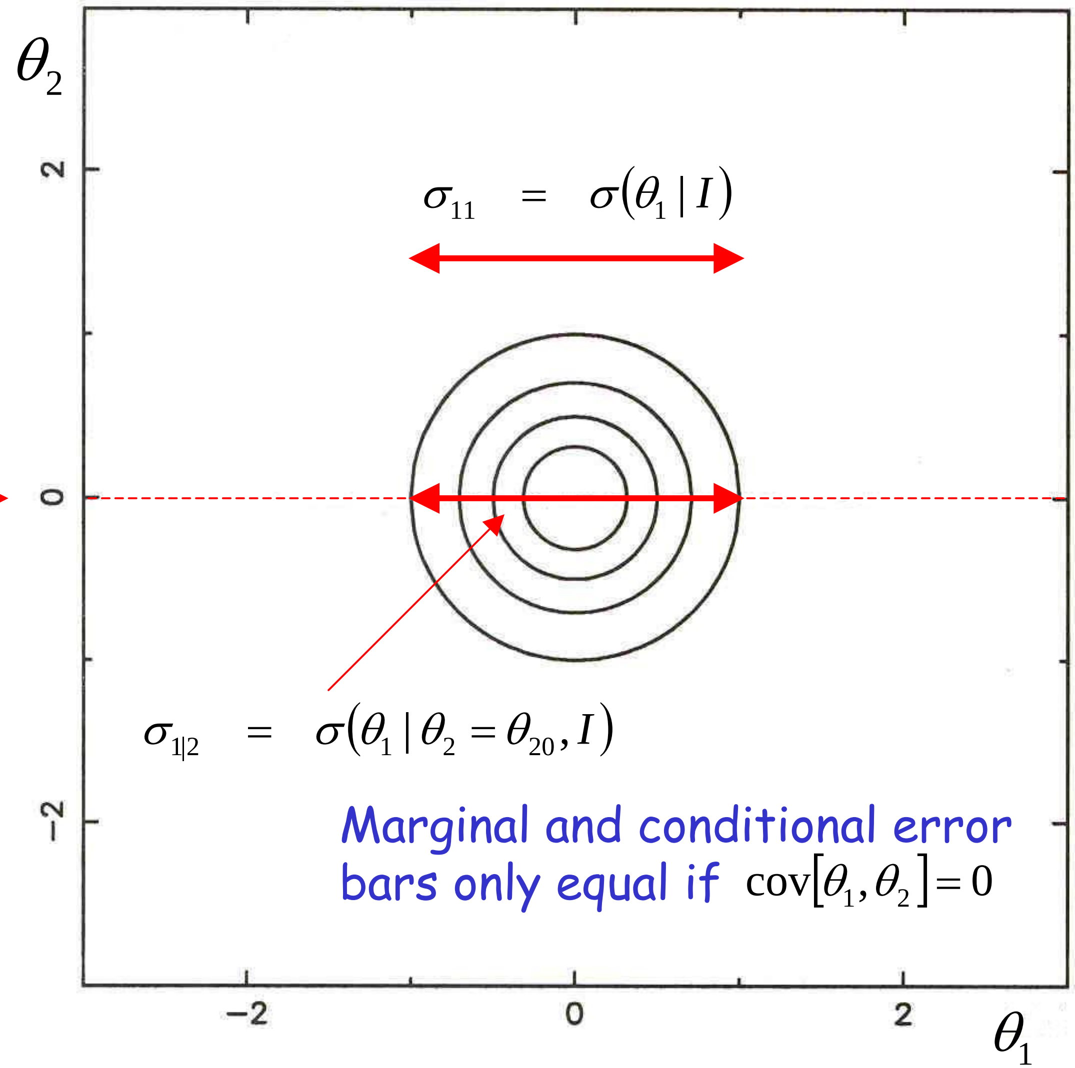
$\Delta\chi^2$ as a Function of Confidence Level and Degrees of Freedom						
	ν					
p	1	2	3	4	5	6
68.3%	1.00	2.30	3.53	4.72	5.89	7.04
90%	2.71	4.61	6.25	7.78	9.24	10.6
95.4%	4.00	6.17	8.02	9.70	11.3	12.8
99%	6.63	9.21	11.3	13.3	15.1	16.8
99.73%	9.00	11.8	14.2	16.3	18.2	20.1
99.99%	15.1	18.4	21.1	23.5	25.7	27.8

UNIVERSITY of GLASGOW

Parameter estimation: 2-D case

'Best-fit' value of θ_2, found from

$$\left.\frac{\partial \ell}{\partial \theta_j}\right|_{\theta_j=\theta_{0j}} = 0$$



$$\sigma_{11} \;=\; \sigma(\theta_1 \mid I)$$

$$\sigma_{1|2} \;=\; \sigma(\theta_1 \mid \theta_2 = \theta_{20}, I)$$

Marginal and conditional error bars only equal if $\mathrm{cov}[\theta_1, \theta_2] = 0$

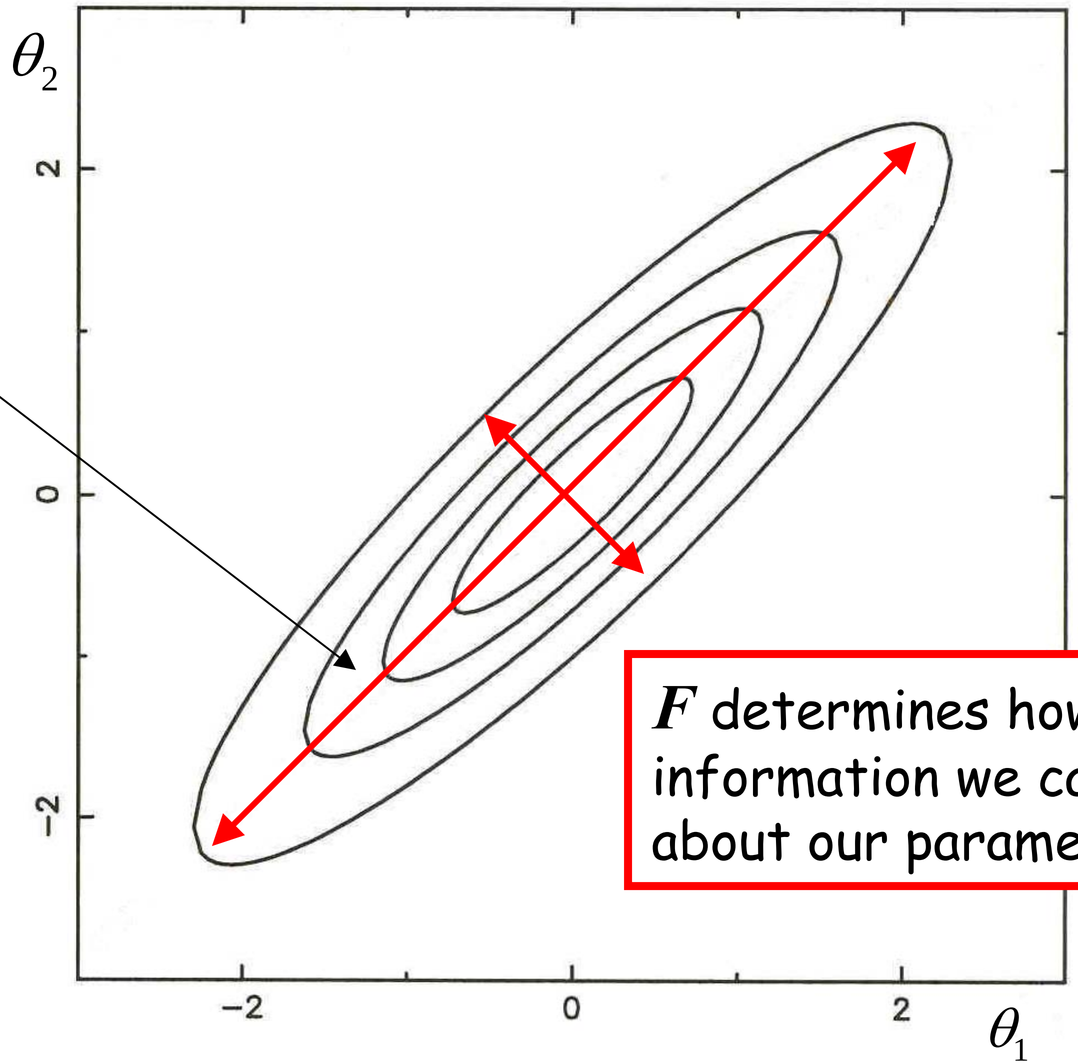
Parameter estimation: 2-D case

Linear combination of θ_1 and θ_2 well constrained by data

Length of axes determined by the **eigenvalues** of the Fisher information matrix

$$F_{ij} = \frac{\partial^2 \ell}{\partial\theta_i \partial\theta_j} = \left[-\sigma_{ij}^2\right]^{-1}$$

$$\boxed{\boldsymbol{F}\,\boldsymbol{\theta} = \lambda\,\boldsymbol{\theta}}$$



F determines how much information we can learn about our parameters

Direction of axes are the *eigenvectors* of ***F***

Parameter estimation: Gaussian approximation

Taylor expand $\ell(\theta_1, \theta_2)$ around θ_{0j} :

$$
\boxed{
\begin{aligned}
\ell(\theta_1,\theta_2) \;=\; & \ell(\theta_{01},\theta_{02}) \;+\; \cancel{\left.\frac{\partial \ell}{\partial \theta_1}\right|_{\theta_j=\theta_{0j}} (\theta_1-\theta_{01})} \;+\; \cancel{\left.\frac{\partial \ell}{\partial \theta_2}\right|_{\theta_j=\theta_{0j}} (\theta_2-\theta_{02})} \;+ \\
& \frac{1}{2}\left[\left.\frac{\partial^2 \ell}{\partial \theta_1^{\,2}}\right|_{\theta_j=\theta_{0j}} (\theta_1-\theta_{01})^2 + \left.\frac{\partial^2 \ell}{\partial \theta_2^{\,2}}\right|_{\theta_j=\theta_{0j}} (\theta_2-\theta_{02})^2 + 2\left.\frac{\partial^2 \ell}{\partial \theta_1 \partial \theta_2}\right|_{\theta_j=\theta_{0j}} (\theta_1-\theta_{01})(\theta_2-\theta_{02}) \right] + \ldots
\end{aligned}
}
$$

$$p(\theta_1, \theta_2 \mid \text{data}, \; I) \;=\; \exp\left[\ell(\theta_1, \theta_2)\right]$$

$$= \exp\left[-\tfrac{1}{2} Q\right]$$

← Gaussian approximation

Is the Gaussian approximation a good idea?

UNIVERSITY of GLASGOW

Parameter estimation: Gaussian approximation

Taylor expand $\ell(\theta_1,\theta_2)$ around θ_{0j} :

$$\ell(\theta_1,\theta_2) = \ell(\theta_{01},\theta_{02}) + \cancel{\left.\frac{\partial \ell}{\partial \theta_1}\right|_{\theta_j=\theta_{0j}} (\theta_1-\theta_{01})} + \cancel{\left.\frac{\partial \ell}{\partial \theta_2}\right|_{\theta_j=\theta_{0j}} (\theta_2-\theta_{02})} +$$

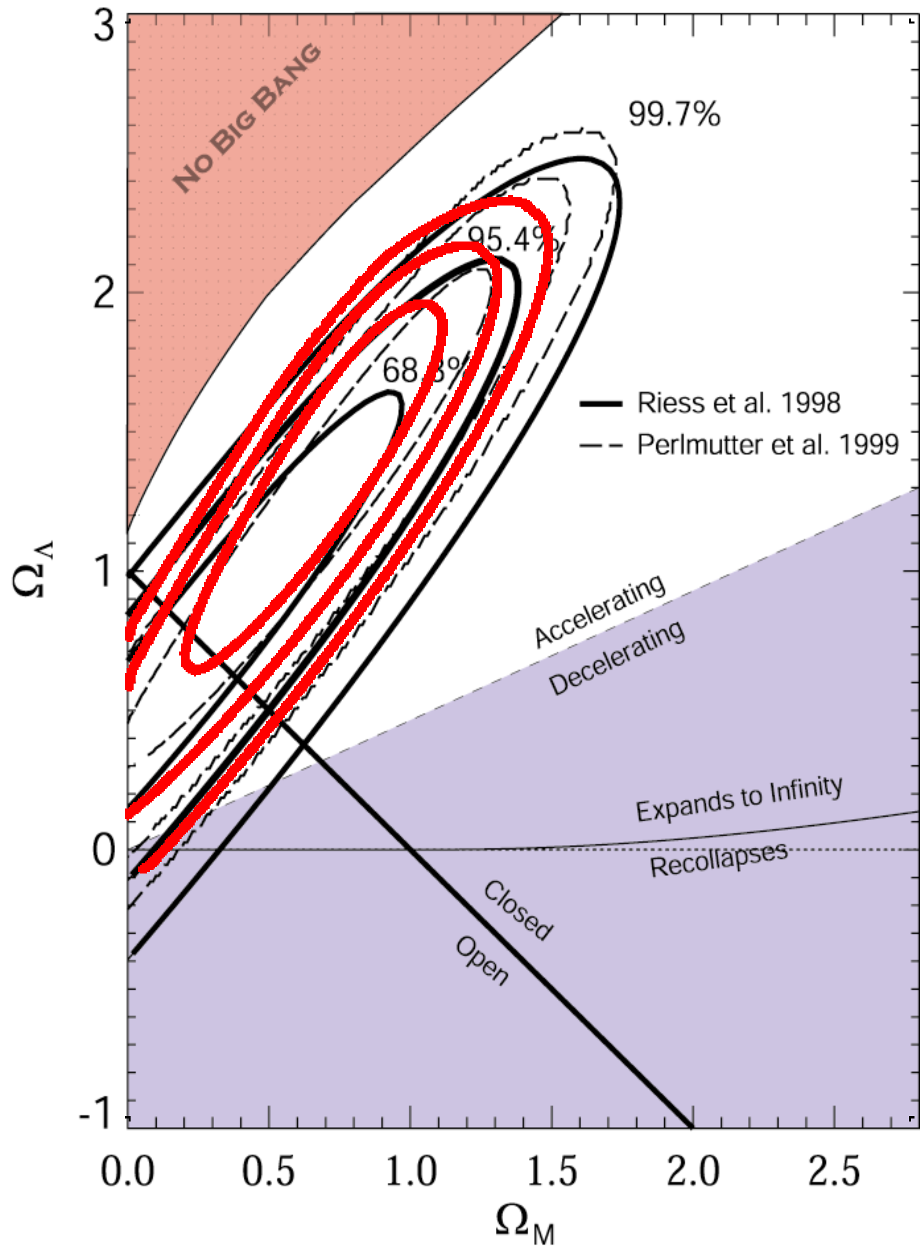
$$\frac{1}{2}\left[\left.\frac{\partial^2 \ell}{\partial \theta_1^{\,2}}\right|_{\theta_j=\theta_{0j}} (\theta_1-\theta_{01})^2 + \left.\frac{\partial^2 \ell}{\partial \theta_2^{\,2}}\right|_{\theta_j=\theta_{0j}} (\theta_2-\theta_{02})^2 + 2\left.\frac{\partial^2 \ell}{\partial \theta_1 \partial \theta_2}\right|_{\theta_j=\theta_{0j}} (\theta_1-\theta_{01})(\theta_2-\theta_{02}) \right] + \dots$$

Is the Gaussian approximation a good idea?

- Greatly simplifies calculations – only need to compute the elements of the Fisher matrix (covariance matrix)
- Nowadays much better to compute **full posterior** pdf. Not too hard with present-day computers, even for large N

UNIVERSITY of GLASGOW

Markov Chain Monte Carlo Methods



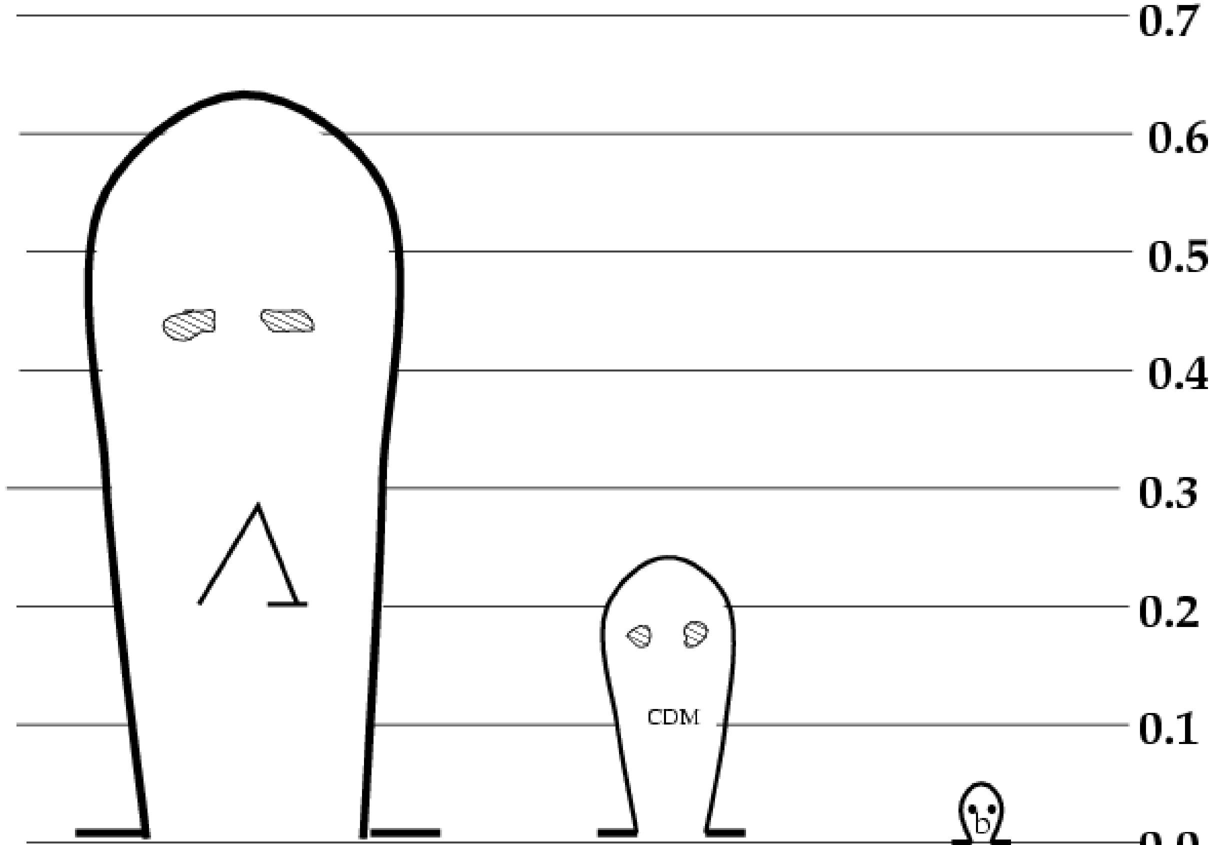
UNIVERSITY of GLASGOW

Λ CDM

Figure 3. A line up of cosmological culprits Ω_Λ is the big shot controlling the Universe. He's going to make it blow up. Ω_{CDM} would like to make the Universe collapse but can't compete with Ω_Λ. Ω_b just follows Ω_{CDM} around. Like all dangerous criminals, one can never be sure of Ω_Λ until he is behind bars. The CMB police is being beefed up. Hundreds of heroic CMB observers are now planning his capture.

From Lineweaver (1998)

Cosmology's Most Wanted



Ω_Λ	Ω_{CDM}	Ω_b
cosmological constant energy of the vacuum He never clumps His evil plan is to blow up the Universe	cold dark matter He likes to clump but has never been detected directly His evil plan is to make the Universe collapse	normal baryonic matter a pawn in the cosmic game who just follows CDM around. He thinks he's a complex life form but is really just a bunch of hydrogen

General Relativity:-

Geometry $\longleftrightarrow$ matter / energy

> *"Spacetime tells matter how to move and matter tells spacetime how to curve"*

Einstein's Field Equations

$$\underbrace{G^{\mu\nu}}_{\text{Einstein tensor}} = \underbrace{R^{\mu\nu}}_{\text{Ricci tensor}} - \frac{1}{2}\underbrace{g^{\mu\nu}}_{\text{Metric tensor}}\underbrace{R}_{\text{Curvature scalar}} = 8\pi G \underbrace{T^{\mu\nu}}_{\text{Energy-momentum tensor of gravitating mass-energy}}$$

General Relativity:-

Geometry $\longleftrightarrow$ matter / energy

> *"Spacetime tells matter how to move and matter tells spacetime how to curve"*

Einstein's Field Equations

Given $g^{\mu\nu}$ can compute $R^{\mu\nu}$ and R ;

These are *generated* by $T^{\mu\nu}$

UNIVERSITY of GLASGOW

Treat Universe as a *perfect fluid*

$$\boxed{T^{\mu\nu} = (\underbrace{\rho}_{\text{Density}} + \underbrace{P}_{\text{Pressure}})\underbrace{u^{\mu}u^{\nu}}_{\text{Four-velocity}} + Pg^{\mu\nu}}$$

N.B. $c = 1$

Solve to give *Friedmann's Equations*

$$H^2 = \left(\frac{\dot{R}}{R}\right)^2 = \frac{8\pi G\rho}{3} - \frac{k}{R^2}$$

$$\frac{\ddot{R}}{R} = -\frac{4\pi G}{3}(\rho + 3P)$$

UNIVERSITY of GLASGOW

IAU

Einstein originally sought *static solution* i.e. :-

$$\dot{R} = 0 \quad \text{for all} \quad t$$

But if $\rho, P \geq 0$ can't have $\ddot{R} = 0$

However, GR actually gives

Covariant derivative

$$G^{\mu\nu}{}_{;\nu} = T^{\mu\nu}{}_{;\nu} = 0$$

Can add a constant times $g^{\mu\nu}$ to $G^{\mu\nu}$

Einstein's cosmological constant

$$\boxed{G^{\mu\nu} = R^{\mu\nu} - \frac{1}{2} g^{\mu\nu} R + g^{\mu\nu} \Lambda}$$

UNIVERSITY of GLASGOW

Friedmann's Equations now give:-

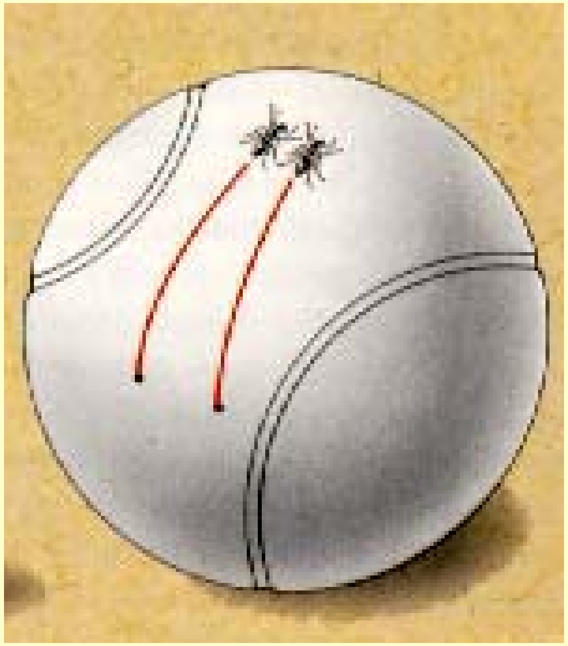
$$H^2 = \left(\frac{\dot{R}}{R}\right)^2 = \frac{8\pi G\rho}{3} + \frac{\Lambda}{3} - \frac{k}{R^2}$$

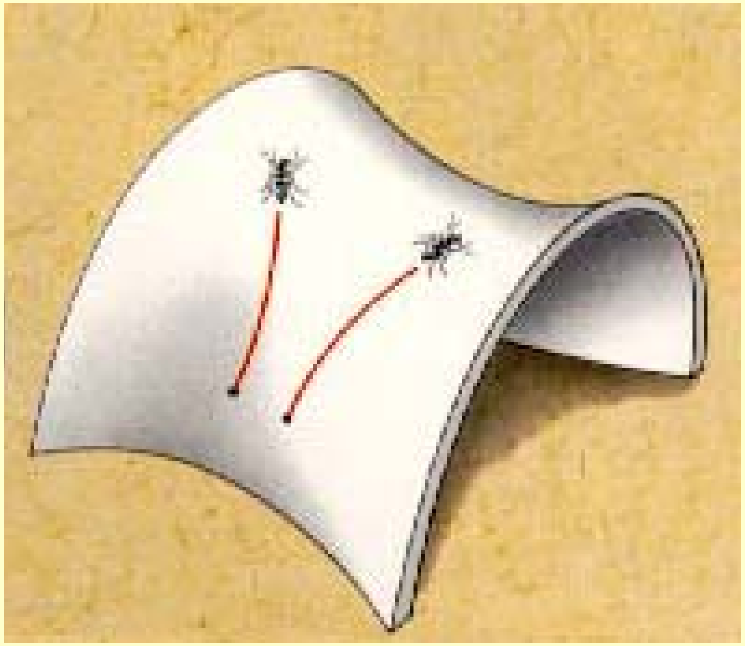
$$\frac{\ddot{R}}{R} = -\frac{4\pi G}{3}\left(\rho + 3P\right) + \frac{\Lambda}{3}$$

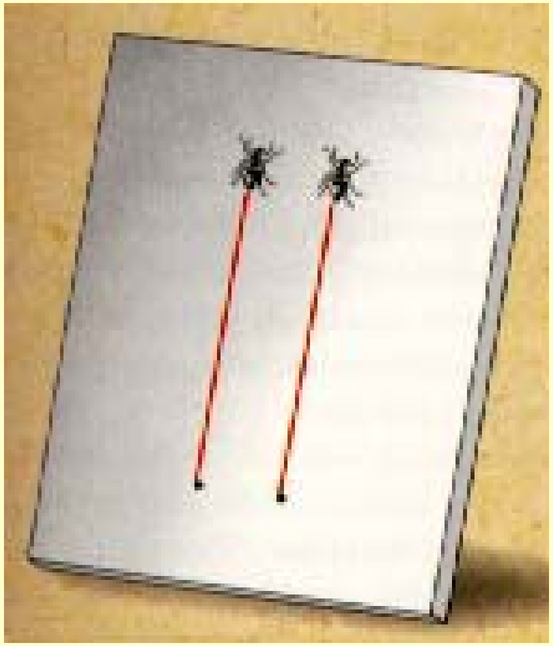
Can tune Λ to give $\dot{R} = 0 \text{ for all } t$ but *unstable*

(and *Hubble expansion* made idea redundant)

But Lambda term could still be non-zero anyway !


Closed


Open


Flat

$$k = \text{curvature constant} = \begin{cases} -1, & \text{open} \\ 0, & \text{flat} \\ +1, & \text{closed} \end{cases}$$

Einstein's greatest blunder?

Re-expressing Friedmann's Equations:-

For $\Lambda = 0$

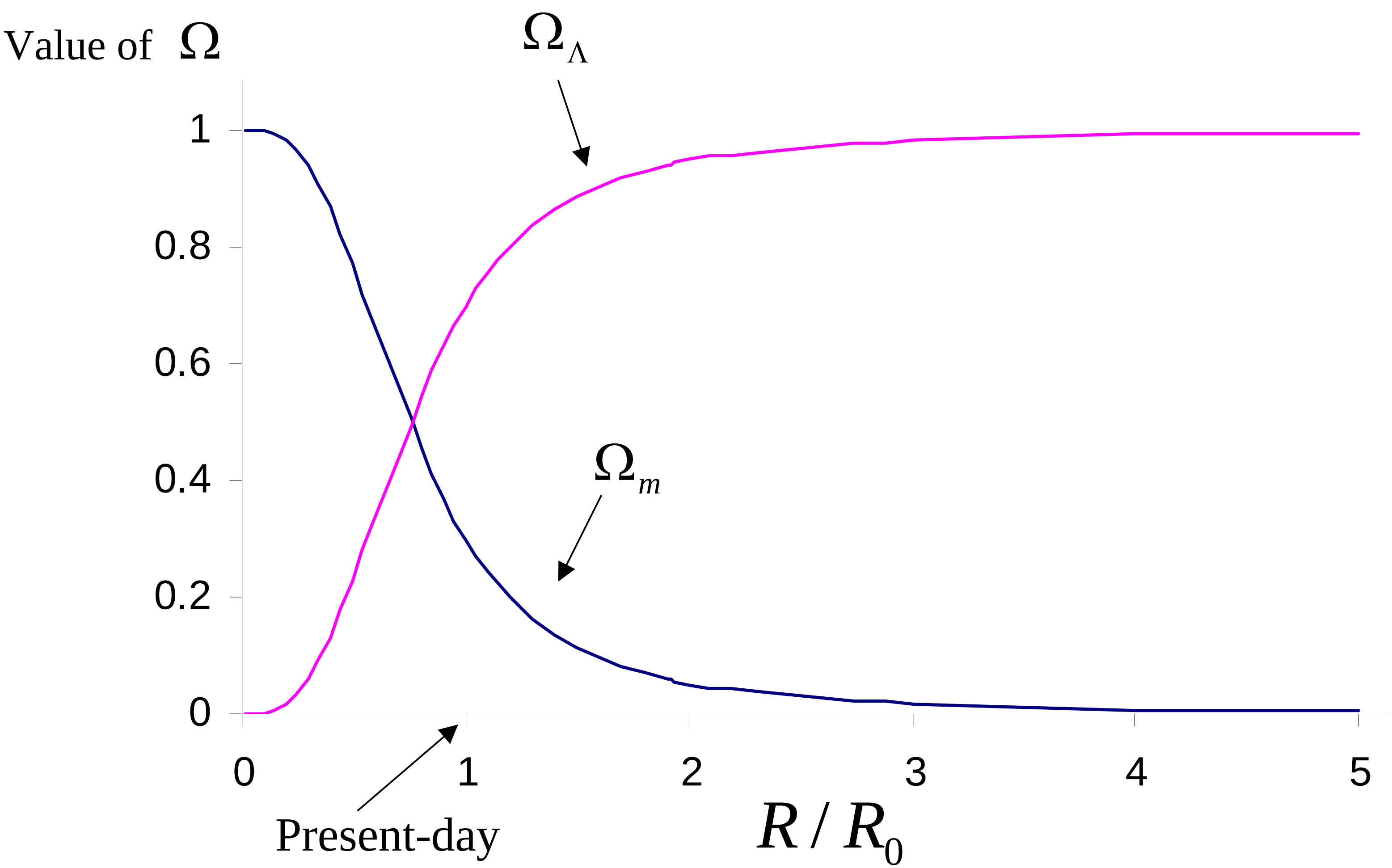
$$H^2 = \frac{8\pi G\rho}{3} - \frac{k}{R^2} \quad \Rightarrow \quad k = 0 \Leftrightarrow \rho = \left[\frac{8\pi G}{3H^2}\right]^{-1} = \rho_{\text{crit}}$$

Define

$$\Omega_m = \rho \big/ \rho_{crit} = \frac{8\pi G\rho}{3H^2} \qquad \Omega_\Lambda = \frac{\Lambda}{3H^2} \qquad \Omega_k = -\frac{k}{R^2 H^2}$$

It follows that, at *any time*

$$\boxed{\Omega_m + \Omega_\Lambda + \Omega_k = 1}$$



If the Concordance Model is right, we live at a special epoch. Why?...

Hubble diagram of distant supernovae

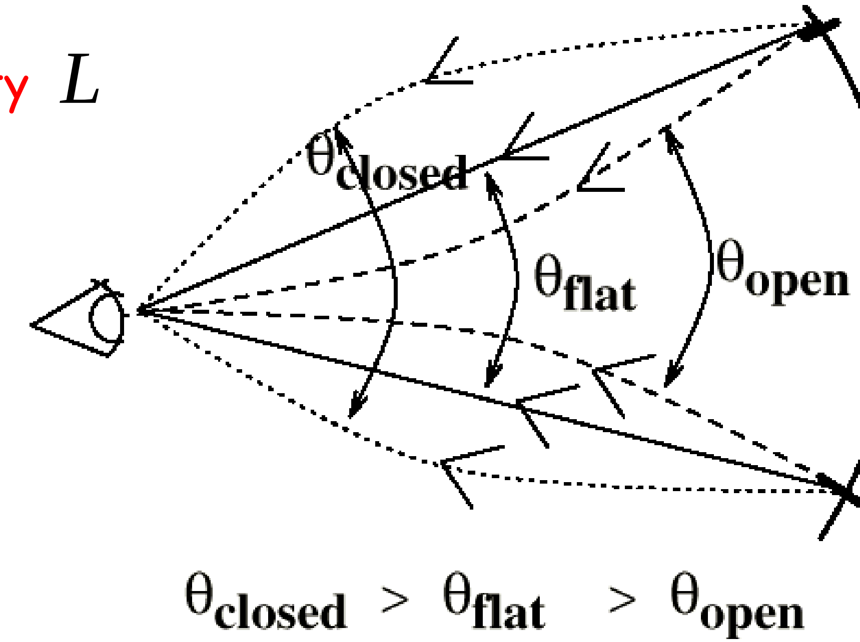
Consider an object of intrinsic luminosity L

from which we observe a flux $\mathfrak{I}$

Define the Luminosity Distance via:-

$$\boxed{\mathfrak{I} = \frac{L}{4\pi d_L^{\,2}}}$$

Distance required to give observed flux *if* Universe has a flat geometry



Actual distance depends on true geometry, and expansion history of the Universe

$$\boxed{d_{\mathrm{L}} \equiv d_{\mathrm{L}}(z;\Omega_m,\Omega_\Lambda) = (1+z)^2 d_{\mathrm{ang}}(z;\Omega_m,\Omega_\Lambda)}$$

Adapted from Schmidt (2002)

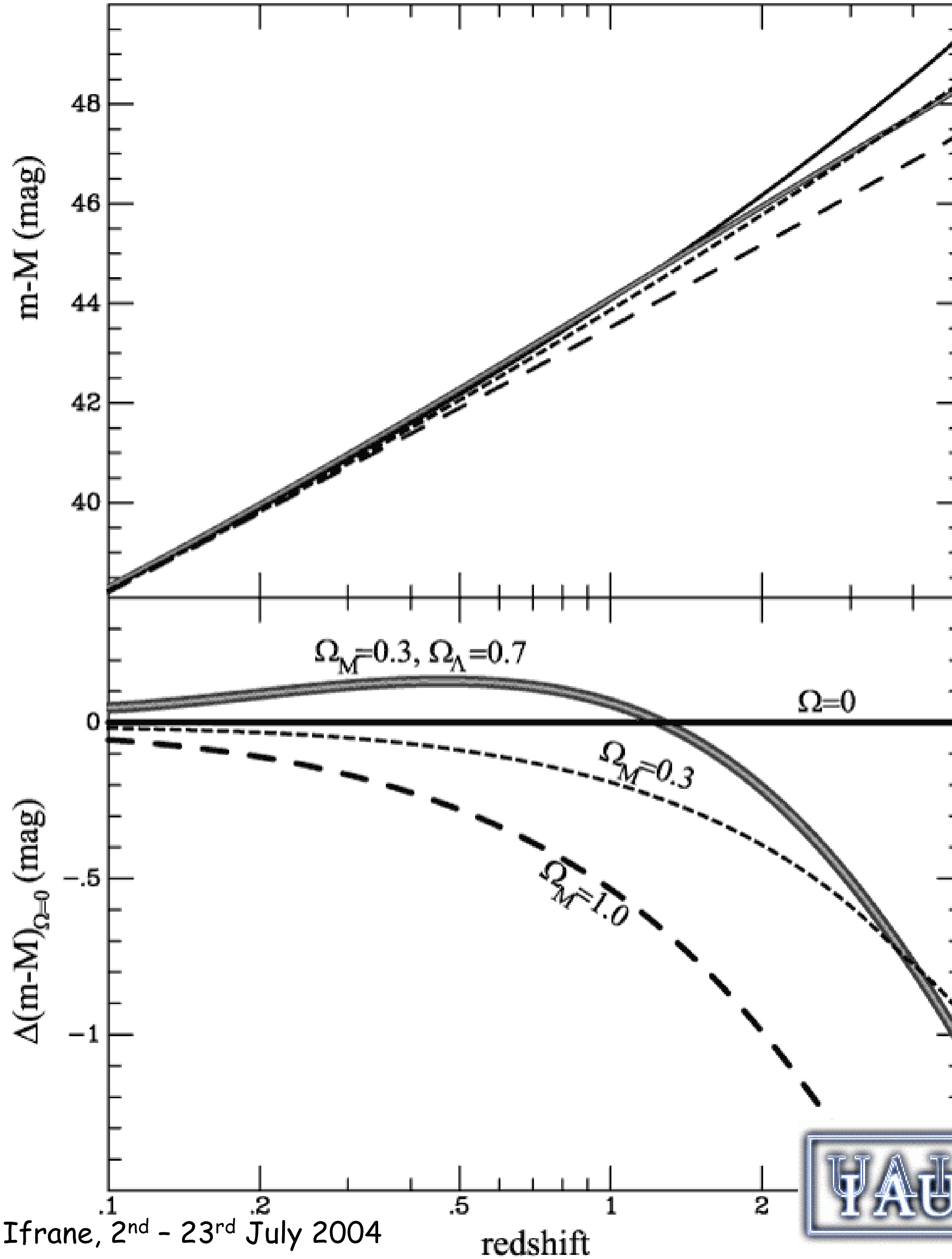
Distance Modulus

$$(m - M)_{mag} = 5\log\left[\frac{d_L}{Mpc}\right] + 25$$

Fractional distance change
$\cong \frac{1}{2}$(mag change)

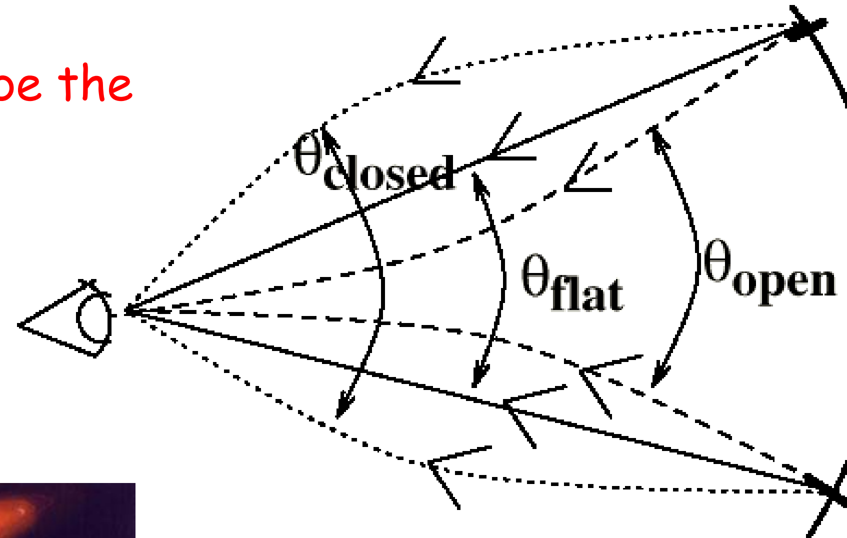
e.g.

0.1 mag difference is 5% distance difference



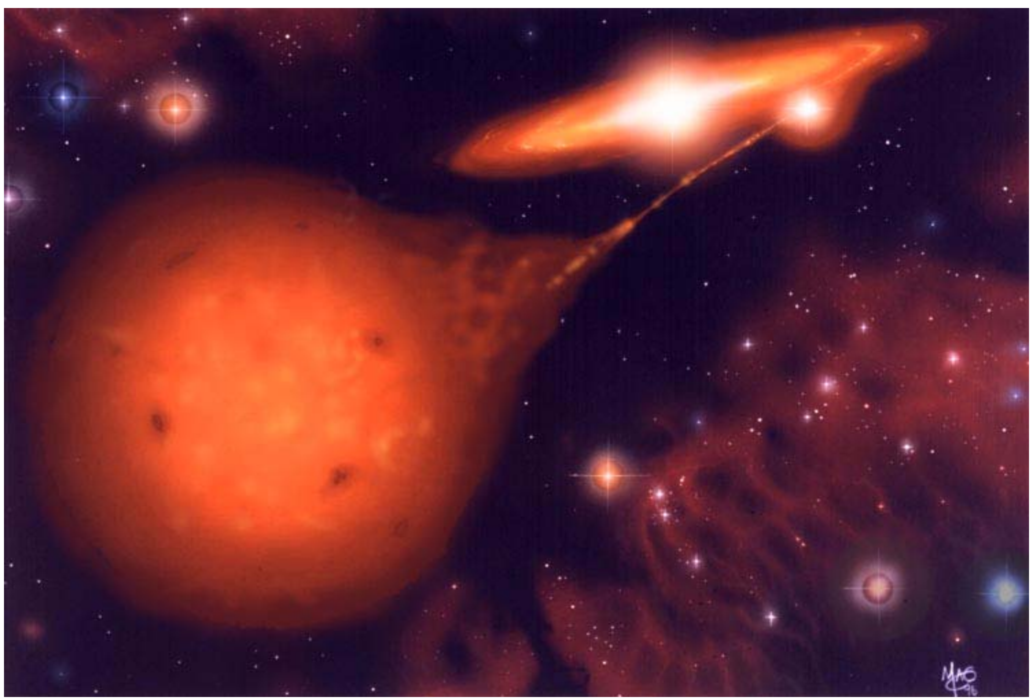
Hubble diagram of distant supernovae

We need a good standard candle, to probe the geometry of the Universe



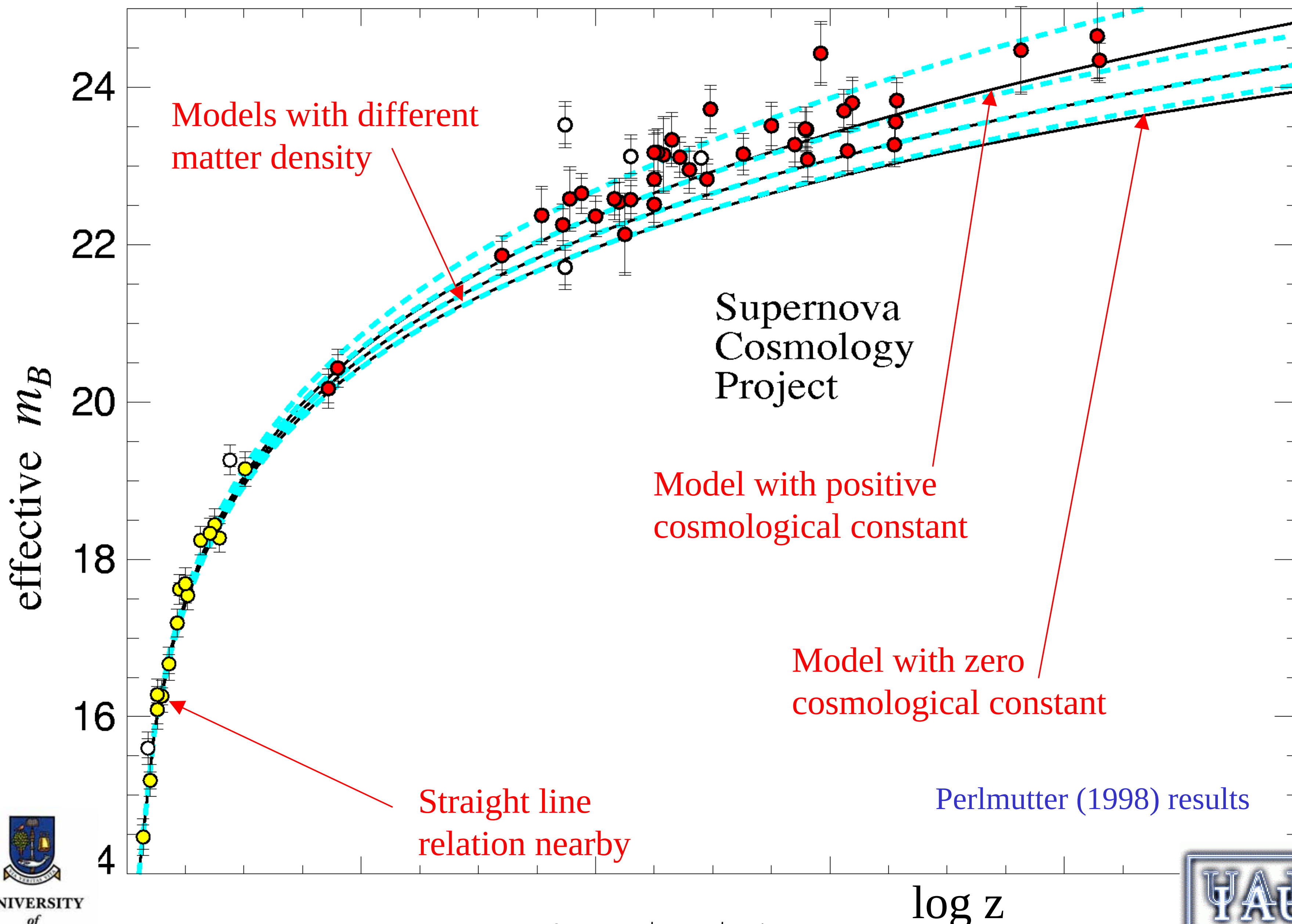
$\theta_{closed} > \theta_{flat} > \theta_{open}$

Type Ia Supernovae



- Visible to huge distances
- Small scatter in luminosity at peak brightness
- Observational programs to detect dozens (100s)
- Can also use `Light Curve Shape' to improve distance estimates

Hubble diagram of distant Type Ia supernovae



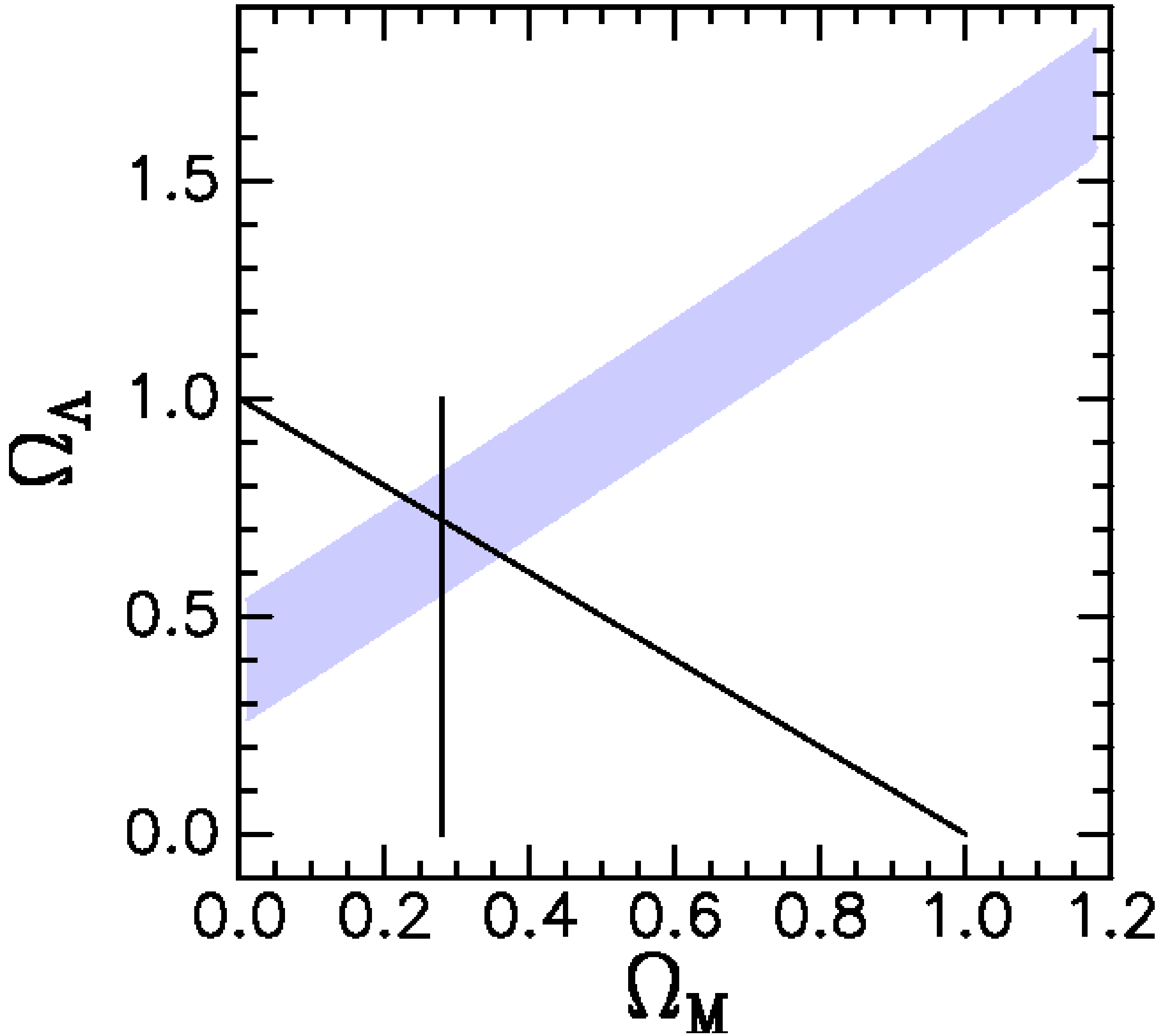
SNIa at z = 0.5

At low redshift, SN1a essentially measure the deceleration parameter

$$q_0 \equiv \frac{-R(t_0)R(t_0)}{\dot{R}^2(t_0)}$$

$$= \frac{1}{2}\sum_i \Omega_i (1+3w_i)$$

$$e.g.,\ q_0 = \frac{1}{2}\Omega_m - \Omega_\Lambda$$



Adapted from Schmidt (2002)

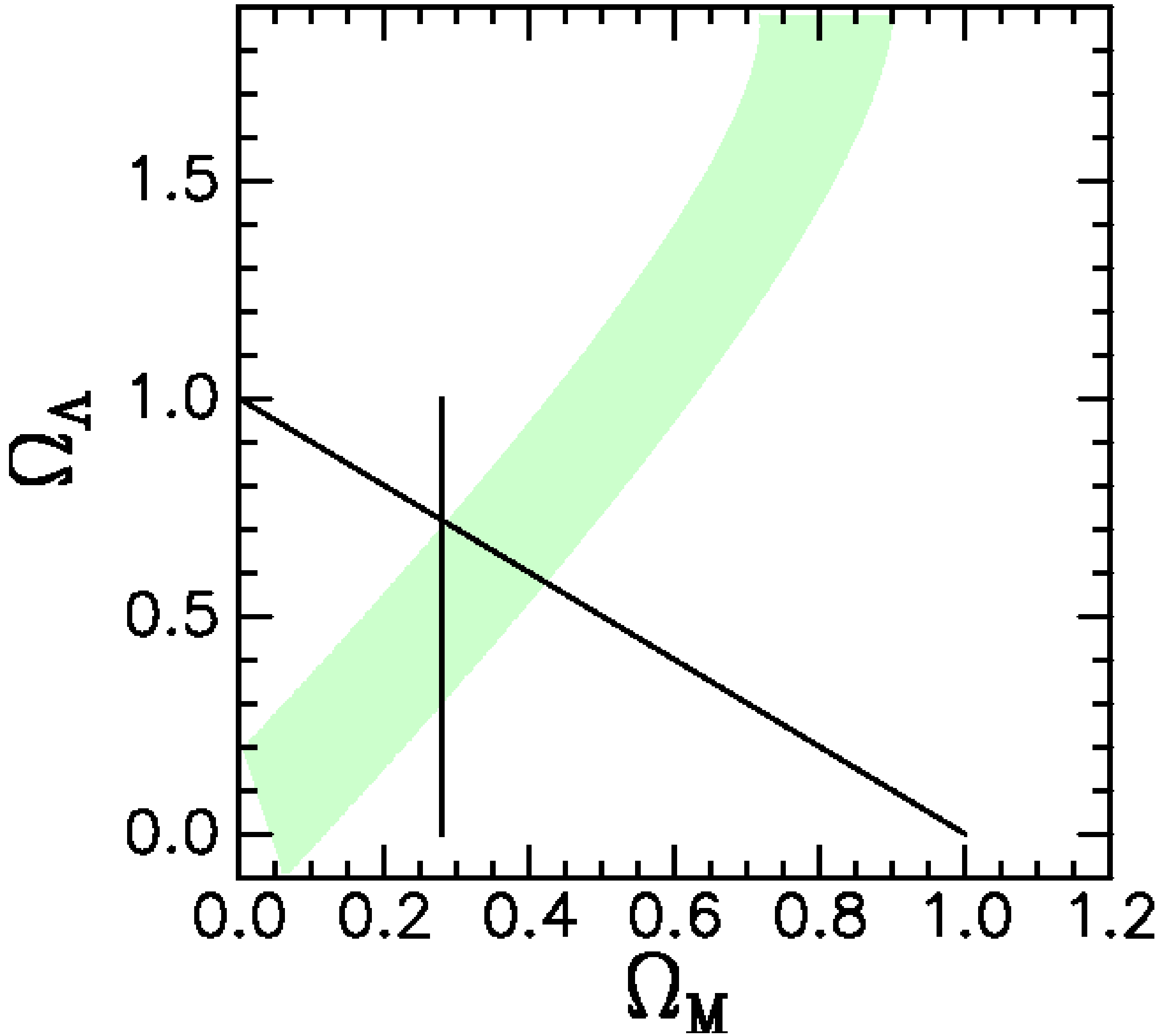
SNIa at z = 1.0

At low redshift, SN1a essentially measure the deceleration parameter

$$q_0 \equiv \frac{-\ddot{R}(t_0)R(t_0)}{\dot{R}^2(t_0)}$$

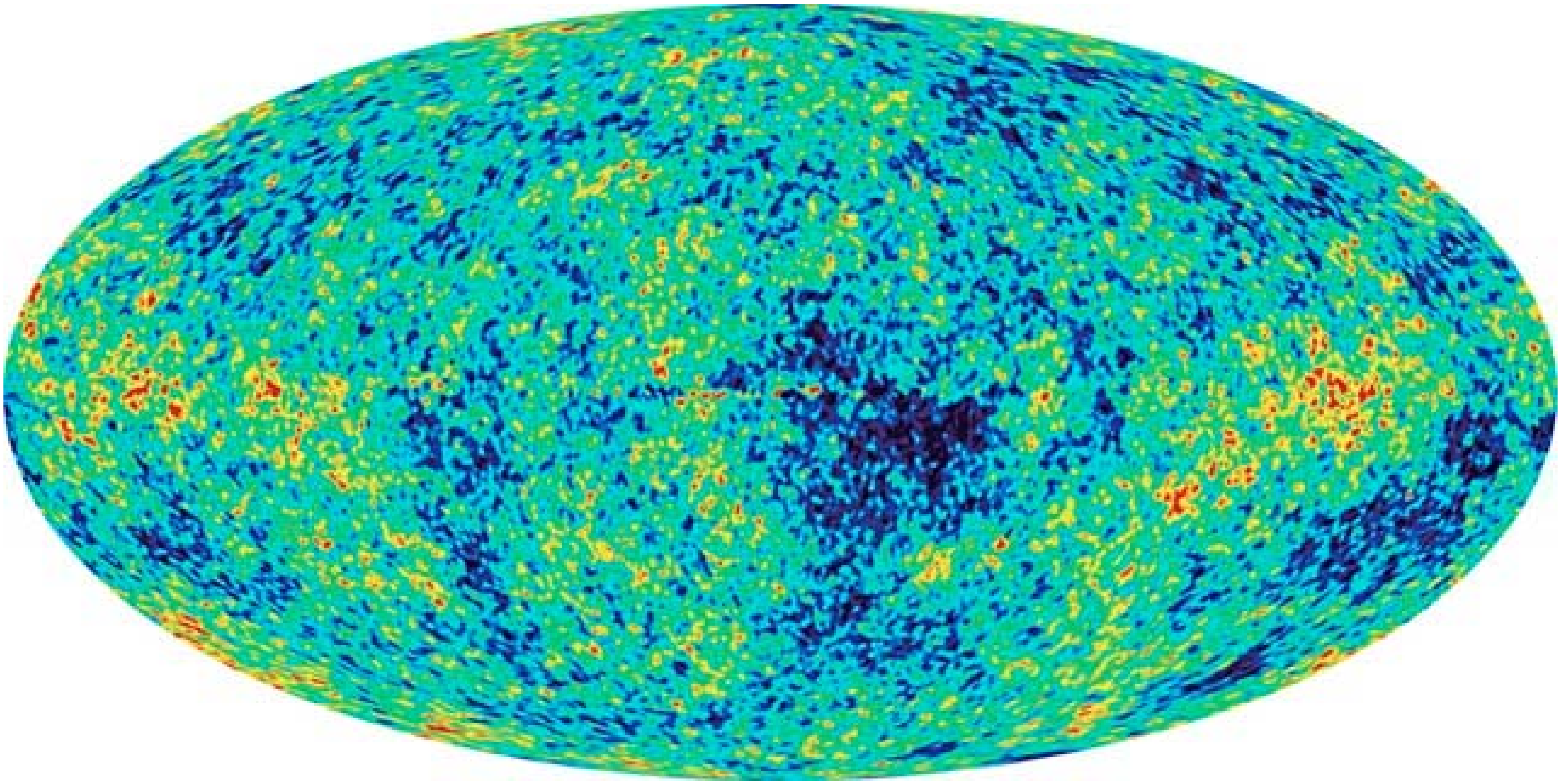
$$= \frac{1}{2}\sum_i \Omega_i (1+3w_i)$$

$$e.g.,\ q_0 = \frac{1}{2}\Omega_m - \Omega_\Lambda$$



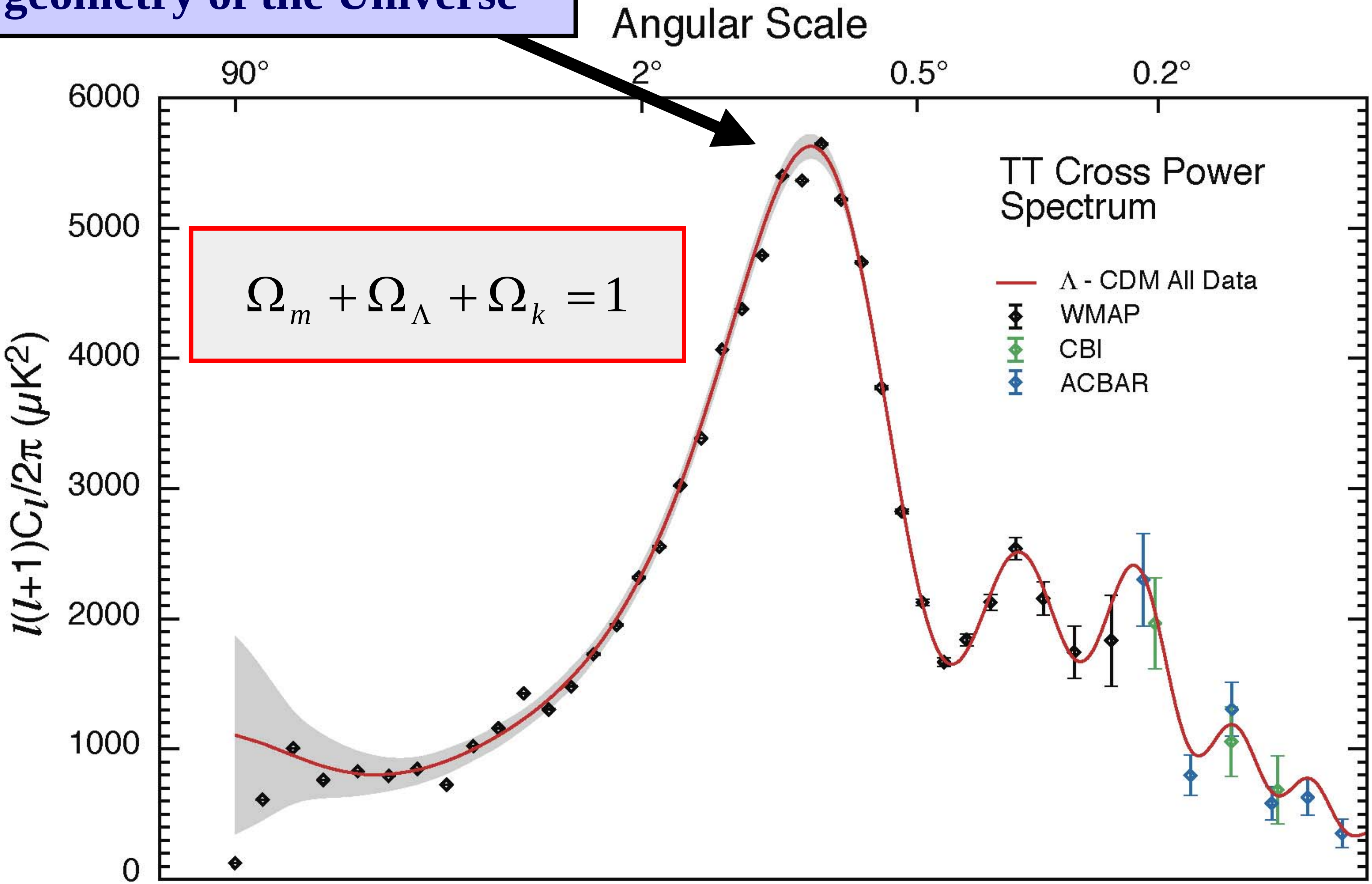
Adapted from Schmidt (2002)

From Bennett et al (2003)



UNIVERSITY
of
GLASGOW

Position of first peak sensitive probe of the geometry of the Universe



$$\Omega_m + \Omega_\Lambda + \Omega_k = 1$$

Tegmark et al (1998)

SNIa measure:-

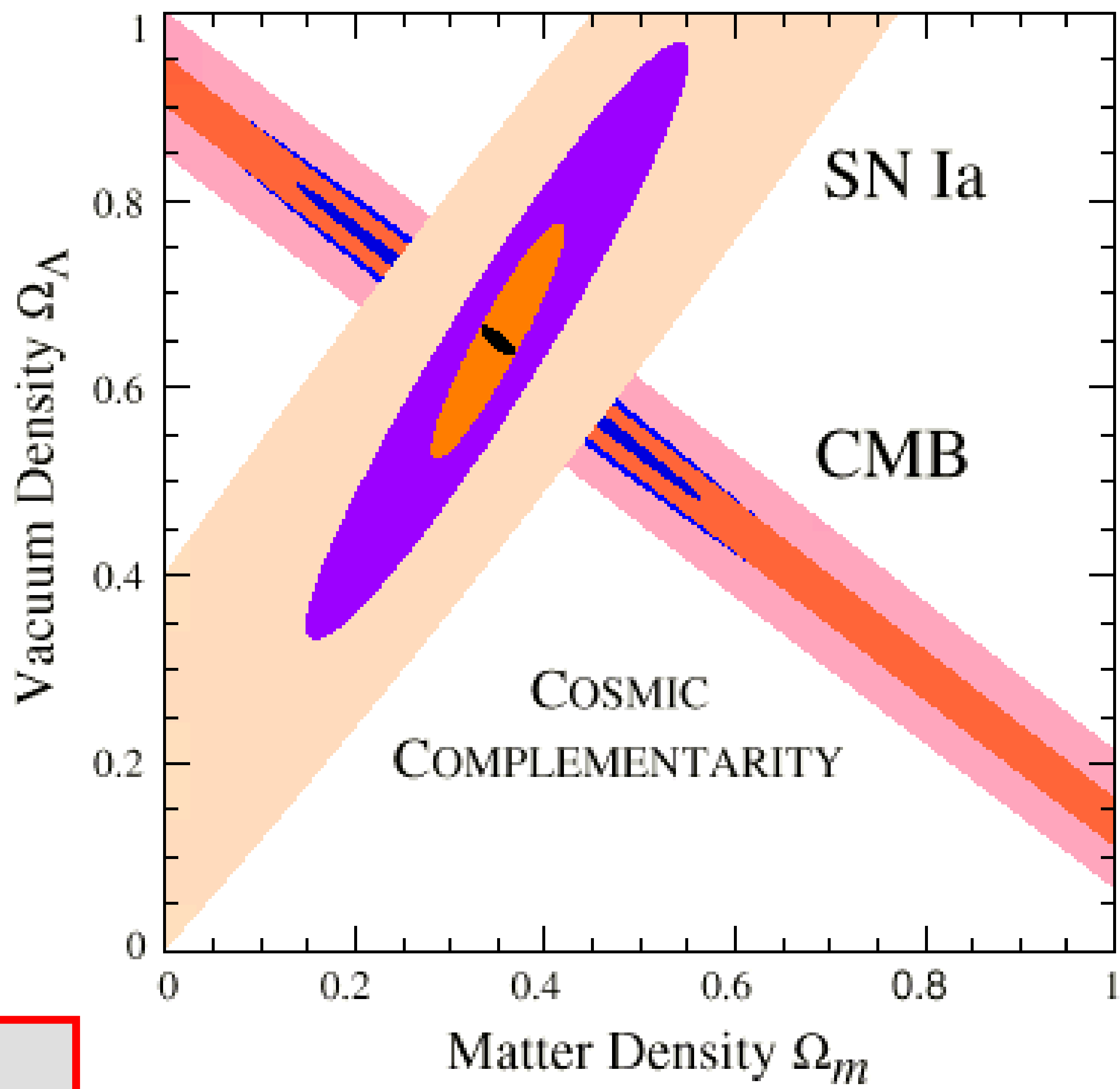
$$q_0 = \frac{1}{2}\Omega_m - \Omega_\Lambda$$

CMBR measures:-

$$\Omega_k = 1 - \Omega_m - \Omega_\Lambda$$

Together, can constrain:-

$$\boxed{\Omega_m, \Omega_\Lambda}$$

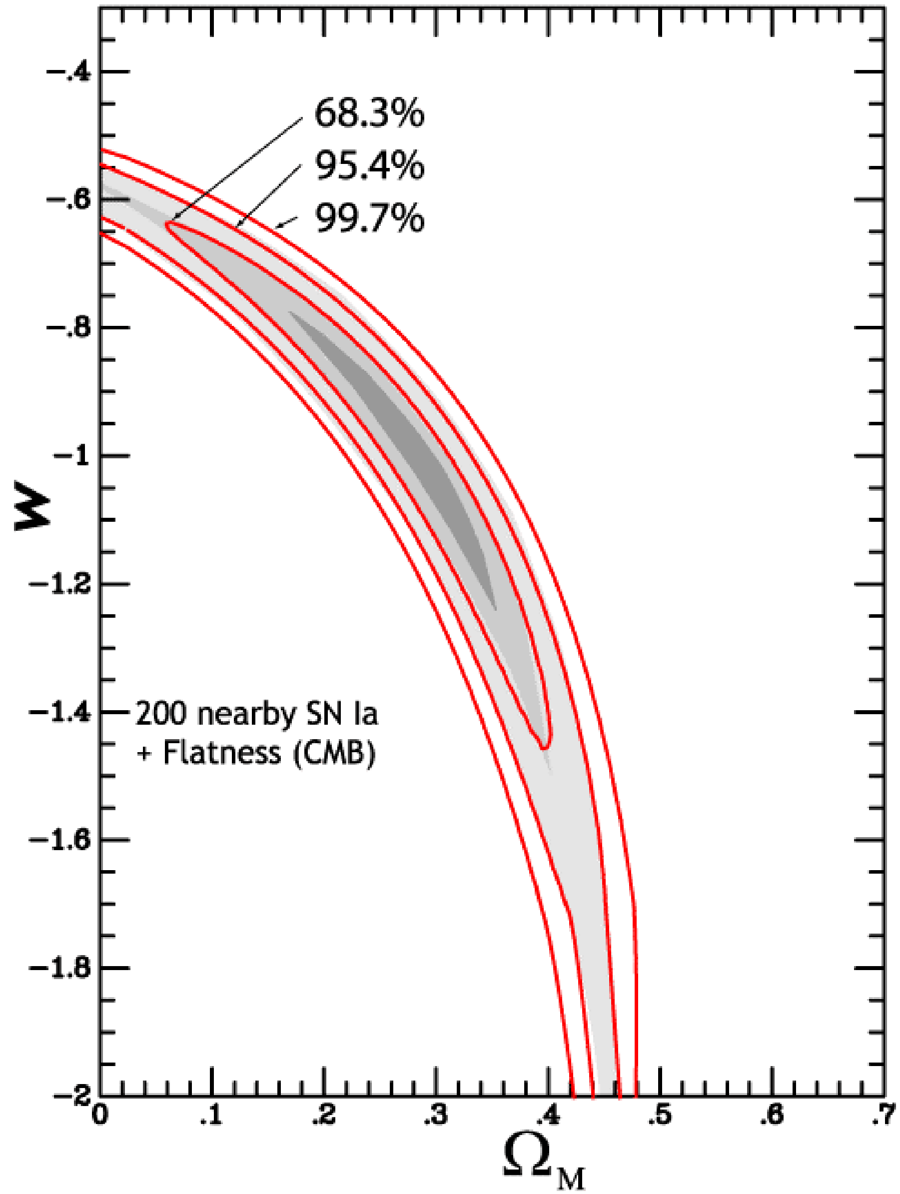


UNIVERSITY of GLASGOW

Can we distinguish a constant Λ term from quintessence?…

Not from current ground-based SN observations (combined with e.g. LSS)…

…or from future ground-based observations

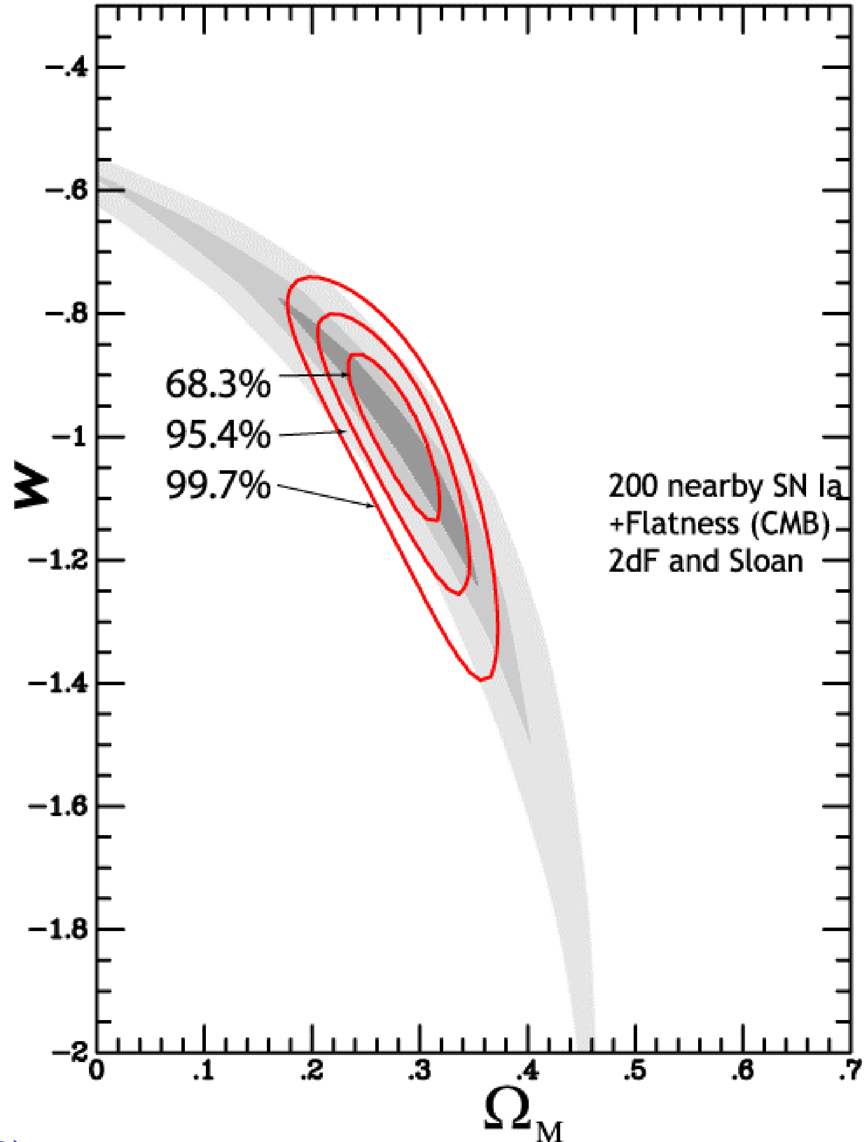


Adapted from Schmidt (2002)

Can we distinguish a constant Λ term from quintessence?...

Not from current ground-based SN observations (combined with e.g. LSS)...

...or from future ground-based observations



Adapted from Schmidt (2002)

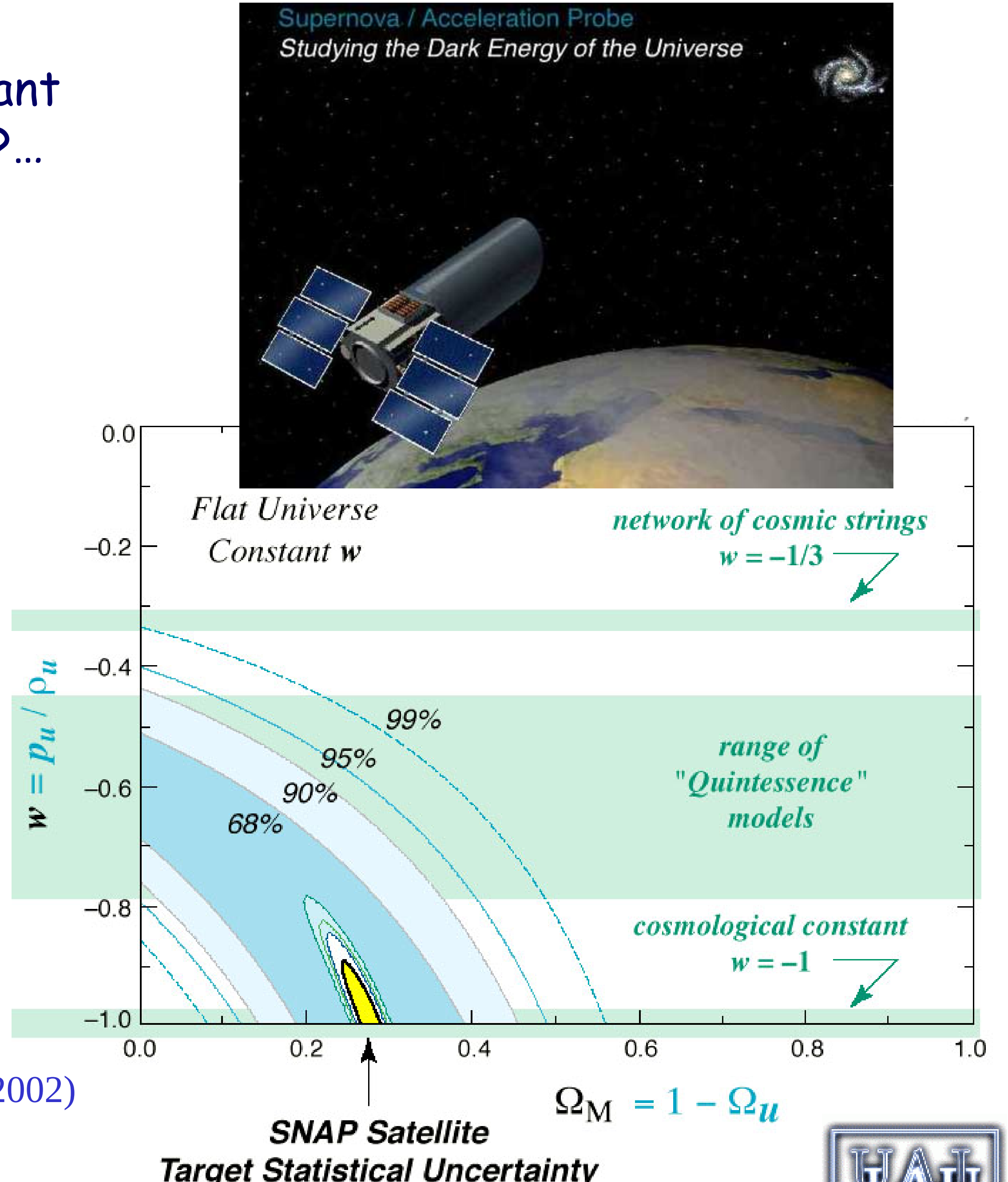
UNIVERSITY of GLASGOW

Can we distinguish a constant Λ term from quintessence?…

Not from current ground-based SN observations (combined with e.g. LSS)…

…or from future ground-based observations

Main goal of the SNAP satellite (launch ~2010?)



Adapted from Schmidt (2002)